

alignment_block:
US-09-768-877-2 x AB058748
Align seg 1/1 to: AB058748 from: 1 to: 2536

1 MetArgAlaGlyArgGlyAlaThrProAlaArgLeuPheArgAspAl 17
101 ATCGGGCGGGCGGGCGGCGAGCGCGCGGAGGAGCTGTTCGGGAGCG 150

17 aAlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrProL 34
151 CGCTTCCCCCGCGCGGACTCTCGCTCTCTCTCGGACTTGTCTACGCCGC 200

34 euAlaGlnPheArgGluAspIleThrTrpArgArgProGlnGluIleCys 50
201 TGGCCCAAGTTCCCGAGGACATCACGTGGAGGGCGGCCACAGAGATTGT 250

51 AlaThrProArgLeuPheProAspAspProArgGluGlnVallysGI 67
251 GCCACACCCCGGCTGTTTCCAGATGACCCACGGGAAGGGGAGGTGAAGCA 300

67 nGlyLeuLeuGlyAspCysTrpPheLeuCysAlaCysAlaAlaLeuGlnL 84
301 GGGGCTGCTGGGGGATTGCTGGTTCTCTGTGCTGCTGGCGCGCGCTGCAGA 350

84 ysSerArgHisLeuLeuAspGlnValIleProProGlyGlnProSerTrp 100
351 AGAGCAGGACCTCTCTGGACAGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 400

101 AlaAspGlnGluThrArgGlySerPheThrCysArgIleTrpGlnPheGI 117
401 GCGGACAGAGTACCGGGGCTCTCTTACCTTGGCATTTGGCAGTTTG 450

117 yArgTrpValGluValThrThrAspAspArgLeuProCysLeuAlaGlyA 134
451 ACGCTGGTGGAGTGACACACAGATGACCGCTGCGCTGCTGCTGCTGCTGCTG 500

134 rgLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
501 GACTCTGTTTCTCCCTGCCAGAGGAGGATGTCTCTGCTGCTGCTGCTGCTG 550

151 LeuGluLysValTyrrAlaLysValHisGlySerTyrrGluHisLeuTrpAl 167
551 CTGGAAGAGGTCTACCCCAAGGTCTATGGTCTCTACGAGACCTGTGGGC 600

167 agGlyGlnValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGluA 184
601 CGGGCAGGTGGCGGATGCCCTGGTGGACCTTGACCGCGGCGCTGGCAGAAA 650

184 rgTrpAsnLeuLysGlyValAlaGlySerGlyGlnGlnAspArgPro 200
651 GATGGAACCTGAAGGCGGTAGCAGGAAGCGGAGCGGAGCGGAGGAGGCGCA 700

201 GlyArgTrpGluHisArgThrCysArgGlnLeuLeuHisLeuLysAspGI 217
701 GCGCGCTGGAGCACAGGACTGTCTGGCAGCTGTCTCCACTGAAGGACCA 750

217 nCysLeuIleSerCysValLeuSerProArgAlaGlyAlaArgGluL 234
751 GTGTCTGTAGTCAGTGTGCTGCTCAGCCCCCAGAGCAGGTGCCCGGAGC 800

234 euGlyCluPheHisAlaPheIleValSerAspLeuArgGluLeuGlnGly 250
801 TGGGGGAGTTCATGCTTTCATTTGCTCGGACCTCGGGAGGTCCAGGGT 850

251 GlnAlaGlyGlnCysIleLeuLeuLeuArgIleGlnAsnProTrpGlyAr 267
851 CAGGCGGGCCAGTGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900

267 gArgCysTrpGlnGlyLeuTrpArgGluGlyGlyGlyTrpSerGlnV 284
901 GCGGTGCTGGCAGGGGCTCTGGAGAGAGGGGGGTGAAGGCTGGAGCGCAGG 950

284 aAlaPalaAlaValAlaSerGluLeuLeuSerGlnLeuGlnGluGlyGlu 300
951 TAGATGCGAGCGGTAGCATCTGAGCTCTCTGCCAGCTCCAGGAGGGGAG 1000

301 PheTrpValGluGluGluGluPheLeuArgGluPheAspGluLeuThrVa 317
1001 TTCTGGTGGAGGAGGAGGAGTTCTCAGGAGTTTTCAGCAGCTCACCGT 1050

317 lGlyTyrrProValThrGluAlaGlyHisLeuGlnSerLeuTyrrThrGluA 334
1051 TGCTACCCCGTACGAGGAGCGGCCACCTGCAGAGCCTCTACACAGAGA 1100

334 rgLeuLeuCysHisThrArgAlaLeuProGlyAlaTrpValLysGlyGln 350
1101 GGCTGCTCTCCATACGCGGCGCTGCTGGGCGCTGGGTCAAGGGCCAG 1150

351 SerAlaGlyGlyCysArgAsnAsnSerGlyPheProSerAsnProLysPh 367
1151 TCAGCAGGAGGCTGCCGGAACAACAGCGGCTTTCACCAACAACCCCAATT 1200

367 eTTPLeuArgValSerGluProSerGluValTyrrIleAlaValLeuGlnA 384
1201 CTGGCTGGGGTCTCAGAACCGAGTGTATCATTTGCCGCTCTGCAGA 1250

384 rgSerArgLeuHisAlaAlaAspTrpAlaGlyArgAlaArgAlaLeuVal 400
1251 GATCCAGGCTGCACGCGGCGACTGGGCGCGCGCGCGCGCGCGCTGGTG 1300

401 GlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHisTy 417
1301 GGTGACAGTCATCTTCTGGAGCCAGCGAGCATCCCGGGCAAGCACTA 1350

417 rGlnAlaValGlyLeuHisLeuTrpLysValGluLysArgArgValAsnL 434
1351 CCAGGCTGGGTCTGCACCTCTGGAGGTAGAGAGCGCGGGTCAATC 1400

434 euProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAla 450
1401 TGCTAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1450

451 TyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTyrrTyLe 467
1451 TAGACCGGGAGGTCCACCTGCTGTGAGCTCTCACCGGCTTACTACCT 1500

467 uAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeuLeuA 484
1501 GGCTGCTCCCGCAGCACCCTCTCTGAGAGCGCGCGAGGGAGTTCTCTCTCC 1550

484 rgValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAla 500
1551 GAGTCTTCTTACCGGGCGAGTCTCTCTTACGCCCATCAGGCGAGTGCC 1600

501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyGluTrpGlyThrVa 517
1601 AAGAACACACCCCGGGGAGCCTCTGCTGCGGGGAGTGGGGGAGCCCT 1650

517 lGlnLeuArgGlySerTrpArgValGlyGlnThrAlaGlyGlySerArgA 534
1651 GCAGCTACGGGGTCTTGGAGAGTGGCGAGAGCGCGGGGGGAGCAGGA 1700

534 snPheAlaSerTyrrProThrAsnProCysPheProPheSerValProGlu 550
1701 ACTTTGCTCATACCCCAACCCCTGCTTCTCTCTCTCTCTCTCTCTCTCT 1750

551 GlyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArgPr 567
1751 GCGCTTGGCCCCCGCTGGCTGCGCATCACTCTGCATCAGCAGCTCGCGCC 1800

567 oSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValProG 584
1801 CAGTGACACCGAGTTCACCCCATCGGCTTCCATATCTTCCAGGTCCTCAG 1850

Page 3

```

BASE COUNT      470 a      809 c      886 g      449 t
ORIGIN
CDS
/cclone="MGC:10770 IMAGE:3606959"
/tissue_type="uterus, endometrium adenocarcinoma"
/cclone_lib="NH_MGC_4"
/lab_host="DH10B-R"
/note="vector: pOTB7"
146. . 2164
/codon_start=1
/product="calpain 10"
/protein_id="AA04260.1"
/db_xref="GI:13279050"
/translaton="MRAGCATPARELFRDAAPFADSLFCDLSTPLAOFREDITWR
RPEICLPTPLFPDDPREGQKGLGIDCWFELCALQSLRHLLQVITPGQSPAD
QEQKSGFTCLIMQGRMVEVITDRLPLCLCFQSGQEDVFWPLLEKVTARKNG
SYELMAGTCVADALVDTGLGLERBMNKLKAGVSGGQDQGRNDEHRCRLHLKDDC
LISGCVSLPACAGARELGEHFATFVSDIRELQGGOCILLRLQNPGRCMQGLME
GGEBSOVDAVASSELLSQLQEGEFWEVEEFLEPDELVGPVPEAGHLSLYTER
LICHTRALPAMWQGSAGGCRNNSSGPMKFWLRVSESEVYIAVLQSLRLAAM
AGARARJCLSGTSWSPASIPGKHQVAGLHMKVEKRRNLRPLRSMVPVATACGA
YDEVEHLRCLSPGYTLAVSTFLKDAQPEGLRLRVSTGVSISATIRAVAKNTTPGA
LIPGEGEVTOLRGSMRGOTAGGSRNPASTPVPCCPFSTPEEGRCACVITLHQR
RSPTEFHGICFHTFQVPEGRSGDAPRLQLQELVSCVPHIRAQEVSRLCLLPAGTYK
VVESTILPDEGAFVYIATRIDRPSIHSGEMLGQFQJQEVSVAAVMT"
```

```

alignment_scores:      Quality: 3649.00      Length: 672
                       Ratio: 5.430      Gaps: 0
Percent Similarity: 100.000      Percent Identity: 100.0000

alignment_block:
US-03-766-877-2 x BC004260      ..

Align seg 1/1      to: BC004260      from: 1      to: 2614

```

GC 245

GT 295

Ys 50

GC 245

OL 34
11

```

|||||
596 CTGAAAGGTTCTACGCCAGGTCATGGGTCTCTACAGCACCTGTGGGC 645
|||||
167 aclyGlnValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGluA 184
|||||
646 CGGCAGGTGGCGATCCCTGGTGGACCTTGACCGCGCCCTGGCAGAA 695
|||||
184 rGTPAsnLeuLysGlyValAlaGlySerGlyGlyGlnGlnAspArgPro 200
|||||
696 GATGGAACCTGAAGGGCGTAGCAGGAAGCGAGGCCAGCAGGAGGCCA 745
|||||
201 GlyArgTPGluHisArgThrCysArgGlnLeuLeuHisLeuLysaspG1 217
|||||
746 GGCCGCTGGGAGCACAGGACTTGTGGCAGCTGCTCCACCTGAAGGACCA 795
|||||
217 nCysLeuIleSerCysCysValLeuSerProArgAlaGlyAlaArgGluL 234
|||||
796 GTGTCTCATCAGCTGCTGCTGCTCAGCCACAGCAGGTGCCGGGAGC 845
|||||
234 euGlyGluPheHisAlaPheIleValSerAspLeuArgGluLeuGlnGly 250
|||||
846 TGGGGAGTTCCATGCCCTTCTATGCTCGGACCTGCGGGAGCTCCAGGT 895
|||||
251 GlnAlaGlyGlnCysIleLeuLeuLeuLeuArgIleGlnAsnProTPGlyAr 267
|||||
896 CAGCGGGCCAGTGATCTCTGCTGCTGCGGATCCAGAACCCCTGGGGCCG 945
|||||
267 gArGcysTPGlnGlyLeuTPArgGluGlyGlyGluGlyTPSerGlnV 284
|||||
946 CGGTGCTGCAGGGGCTCTGGAGAGGGGGGTGAAGGTGGAGCCAGG 995
|||||
284 alAspAlaAlaValAlaSerGluLeuLeuSerGlnLeuGlnGluGlyGlu 300
|||||
996 TAGATGACGGGTAGCATCTGAGCTCTGCTGCCAGCTCCAGGAAGGGAG 1045
|||||
301 PheTPValGluGluGluPheLeuAtqGluPheAspGluLeuThrVa 317
|||||
1046 TTCGGTGGAGGAGGAGGAGTTCCTCAGGGAGTTTGACGAGCTCACCGT 1095
|||||
317 lGlyTPProValThrGluAlaGlyHisLeuGlnSerLeuTPThrGluA 334
|||||
1096 TGGCTACCCGGTCACGGAGGCGGCCACCTGCAGAGCCTCTACACAGAGA 1145
|||||
334 rGLeuLeuCysHisThrArgAlaLeuProGlyAlaTPValLysGlyGln 350
|||||
1146 GGTGCTCTCCATACGGGGCGCTGCTGGGGCTGGGTCAAGGGCCAG 1195
|||||
351 SerAlaGlyCysArgAsnAsnSerGlyPheProSerAsnProLysph 367
|||||
1196 TCAGCAGGAGGCTGCCGAACAACAGCGGCTTTCCAGCAACCCCAAAAT 1245
|||||
367 eTPLeuArgValSerGluProSerGluValTyrlleAlaValLeuGlnA 384
|||||
1246 CTGGCTCGGGTCTCAGAACCGAGTCAGGTGTACATTGGCGTCTGCAGA 1295
|||||
384 rGSerArgLeuHisAlaAlaAspTPAlaGlyArgAlaArgAlaLeuVal 400
|||||
1296 GATCCAGGCTGCACGGGGGAGTGGGCAGCGCGCGCGGCTGGTG 1345
|||||
401 GlyAspSerHisThrSerTPSerProAlaSerIleProGlyLysHisTy 417
|||||
1346 GGTGACAGCTCATCTTGTGGAGCCCGCAGGAGCATCCCGGGCAAGCACTA 1395
|||||
417 rGlnAlaValGlyLeuHisLeuTPLysValGluLysArgArgValAsnL 434
|||||
1396 CCAGGCTGTGGGTCTGCACCTCTGGAAGGTAGAGAAGCGGGGTCAATC 1445
|||||
434 euProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAla 450
|||||
1446 TGCTAGGGTCTGTCCATGCCCCCGCTGGCTGGCACCGCTGCCATGCA 1495
|||||
451 TyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTPTyrlle 467
|||||

```

```

1496 TAGACCGCGGAGGTCCACCTCGGTGTGTGAGCTCTCACCGGGCTACTACCT 1545
|||||
467 uAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeuLeuA 484
|||||
1546 GGTGTGTCGCCAGCACCTTCTGAAGGACGCCAGGGGAGTTCTCTGCTCC 1595
|||||
484 rGValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAla 500
|||||
1596 GAGTCTTCTCTACCGGGGAGTCTCCCTTAGCGCCATCAGGCGAGTGCC 1645
|||||
501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyGluTPGlyThrVa 517
|||||
1646 AAGAACACCAACCCCGGGGAGCCCTGCCCTGCGGGGAGTGGGGACCGT 1695
|||||
517 lGlnLeuAtqGlySerTPArgValGlyGlnThrAlaGlyGlySerArgA 534
|||||
1696 GCAGCTACGGGGTCTTGGAGAGTCGCCAGCGGGGGGGGCGAGCAGGA 1745
|||||
534 snPheAlaSerTPProThrAsnProCysPheProPheSerValProGlu 550
|||||
1746 ACTTGGCTCATACCCACCAACCCCTGCTTCCCTTCTCGGTCCCCGAG 1795
|||||
551 GlyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArgPr 567
|||||
1796 GGCCCTGGCCCCGCTGCGTCCGTCATCTGCATCAGCACTGCCGGCC 1845
|||||
567 oSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValProG 584
|||||
1846 CAGTGACACCGAGTTCCACCCATCGCTTCCATATCTTCCAGGTCCCG 1895
|||||
584 luGlyGlyArgSerGlnAspAlaProProLeuLeuLeuGlnProLeu 600
|||||
1896 AGGCTGAGGAGCCAGGAGCGACCCCTGCTGCTGCAGGAGCCGCTG 1945
|||||
601 LeuSerCysValProHisArgTPArgAlaGlnGluValSerArgLeuCysLe 617
|||||
1946 CTGAGTGGCTGCCACATCGCTACGCCCGCAGGAGGTGAGCGCGCTCTGCT 1995
|||||
617 uLeuProAlaGlyThrTyrlsValValProSerTPThrLeuProAspT 634
|||||
1996 CTGCTGCTGAGGACCTACAGGTTGTGCCCTCCACCTACTCTCCGGACA 2045
|||||
634 hrGluGlyAlaPheThrValThrIleAlaThrArgIleAspArgProSer 650
|||||
2046 CAGAGGGGCTTCACAGTGACCATCGCAACCAAGGATTGACAGGCCATCC 2095
|||||
651 IleHisSerGlnGluMetLeuGlyGlnPheLeuGlnGluValSerValMe 667
|||||
2096 ATTCAGCGCAGGAGATGCTGGGCCAGTTCTCTCCAAGAGGTCTCCGTCT 2145
|||||
667 talAlaValMetLysThr 672
|||||
2146 GGCAGTGTGAAACCC 2161
|||||

```

seq_name: gb_pr:BC007553

seq_documentation_block:

LOCUS BC007553 2614 bp mRNA linear PRI 12-JUL-2001
DEFINITION Homo sapiens, clone IMAGE:2967328, mRNA, partial cds.

ACCESSION BC007553

VERSION BC007553.1 GI:14043130

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 2614)

AUTHORS Strausberg, R.

TITLE Direct Submission

JOURNAL Submitted (10-MAY-2001) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

REMARK
COMMENT

NIH-MGC Project URL: <http://mgc.mcl.nih.gov>
Contact: MGC help desk
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC/DCTD/DTF
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@cgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Pandoh, Anna-Jissa Prabhu, Parvaneh Saeedi, Jacqueline
Schein, Duane Smalhus, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranda Tsai, Natasha van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 22 Row: d Column: 20.

FEATURES

source
Location/Qualifiers

CDS

1..2614
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2967328"
/tissue_type="Skin, melanotic melanoma."
/clone_lib="NIH-MGC-20"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
<1..2164
/codon_start=2
/product="Unknown (protein for IMAGE:2967328)"
/protein_id="AA07553.1"
/db_xref="GI:14043131"
/translation="TGGALGAGGAGRNGRGRGPEAAADSPSLRGCDPEATGRMRP
AEPMRAGCATPARRELFRDAAPPAADSSLECDSTPLAOFREDITWRROELCATR
LPDPDPREGOVKGLLDGDFWLCACALQSKRLDLVDVPPGQSMADQYRGSEFTR
IWFGRWVEVTDLDRLPCLAGRLCFCRQREDYDRLLEKVVAKVGSYEHMAGVY
ADALVDLTGGLAERNLKGVAAGSGQDDRPGRMRHRLHLKDCCLSCCCLSPR
AGARELGEFHAFVSDLRLOGAGAGCILLRIQNPGRRCMOGLMRGEGBSQVDA
AVASLLSQOGEFMYEEEFLEFDELTVGTAGHLSLYERLCHRTALRG
AMVKGDSAGCCRNNSGFPNPKFWLRVSESEYIAVLQSRRLAAMAGARALVSD
SHYKSWSPASIPGRHGOAVGLHMKVEKRVNLPRVISMPPVAGTACADYREVLRC
LSPGYLAVPSTFLKDAFGEFLKVEFSTGVSLSAIVAKNTPGCALPAGENGTVG
LRSMSRYGOTAGSRRNFASYPTNDCPEFSVDEGPRCVRITLHQRSPDERHTVG
FHIQVPEGRSDAPRLLOPELSCVPHRYAOEYSRLCLILPAGTKVVPSTLPDT
EGAFVTIARIDRPSIHSQEMLSQFLQEVSYMAVMKT"

BASE COUNT 470 a 809 c 886 g 449 t
ORIGIN

alignment_scores:

Quality: 3649.00 Length: 672
Ratio: 5.430 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-768-877-2 x BC007553 ..

Align seg 1/1 to: BC007553 from: 1 to: 2614

1 Metatgaiaaglyaraglyalathrprolaaarglulaupearaspal 17
|||||
146 ATGG 195
17 aalapherprolaalaaspserserleuhecysaspseuSerThrproL 34
|||||
196 CGCCTTCGGCCCGCGGAGCTCTGCTCTTGGACATTGCTACGGCGC 245
34 euaiaagluhpharagluaspplierhtrtparagargpProglngluLeu 50
|||||
246 TGGCCCACTTCGGAGAGCATCATCGTGGAGCGGGCCCGAGAGATTGCT 295

51 AlaThrProargLeuPheProaspaspproAargglugluValylsGL 67
|||||
296 GCCACACCCCGGCTGTTCCAGATGACCCAGGGAAGCGAGGTGAAGCA 345
67 nglyLeuGlyaspCysstrpPheLeuGlyaspCysalaalaLeuGLn 84
|||||
346 GGGGCTGCGGGGAGATTGCGGTCTCTGTCTGTGCTGCGCGGCTGACA 395
84 ySerargHisLeuLeuaspClnValIleProProglngluInProSertrp 100
|||||
396 AGAGCAGGACCTCTCGACAGCATATTCCTCGGAGACGCGAGCTGG 445
101 AlaaspGluGluThrArgGlySerPheThrCysArgIletrpGlnheG 117
|||||
446 GCCGACAGAGATACCGGGGCTCTTCCCTGCGCATTTGGCAGCTTTGG 495
117 yArtrpValGluValThrThrAspAspArgLeuProCysLeuAlaGlyA 134
|||||
496 AGCGTGGGTGAGGTGACACAGATACCGCTGCGCTTGGTCAAGGA 545
134 rglauCysPheSerArgCysGlnArgGluaspValPheThrLeuProLeu 150
|||||
546 GACTGTGTTTCCCGCTGCGAGAGGAGATGTGTTGGCTCCCTTCA 595
151 leuGluLysValIlyrAlaLysValHisGlySerTyrGluHisLeuTrpAl 167
|||||
596 CTGGAAAGGTCTACGCCAAGGTCCATGGGTCTTACGACACCTGTGGGC 645
167 aclyGluValAlaaspAlaLeuValaspLeuThrGlyGlyLeuAlaGluA 184
|||||
646 CGGGCAGGTGGGGGATGCCCTGGTGAACCTGACCGGGGCTGGCAAGAA 695
184 rGTrpAsnLeuLysGlyValAlaGlySerGlyGlyGlnGlnAspArgPro 200
|||||
696 GATGAACTGAAAGGCGGTACAGCAAGGCAAGCGAGGAGGAGGAGGCA 745
201 GlyArgTrpGluHisArgThrCysArgGlnLeuLeuHisLeuLysAsp 217
|||||
746 GGGCGGTGGAGACACGAGCTTGTGGCAGCTGCTCACCTTAAGGACCA 795
217 nCysLeuIleSerCysCysValLeuSerProArgAlaGlyAlaArgGlu 234
|||||
796 GGTGTGATCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 845
234 euGlyGluPheHisAlaPheIleValSerAspLeuArgGluLeuGlnGly 250
|||||
846 TGGGGAGATTCCATGCTCATTTGCTCGAGCTGGGAGGAGTCCAGGCT 895
251 GlnAlaGlyGlnCysIleLeuLeuLeuArgIleGlnAspTrpGlyArg 267
|||||
896 CAGGGGGGCGAGTGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 945
267 gArGcYstrpGlnGlyLeuTrpArgGluGlyGlyGlnGlyTrpSerGln 284
|||||
946 GGGGTGCTGGAGGGGCTTGGAGAGAGGGGGGTGAAGGTTGAGACCAAG 995
284 aLaspAlaAlaValaLaspSerGluLeuLeuSerGlnLeuGlnGlyGlu 300
|||||
996 TAGATGACAGCGGTACATCTGAGCTCTGCTGCTGCTGCTGCTGCTGCT 1045
301 PheTrpValGluGluGluGluPheLeuArgGluLeuAspGluLeuThrVa 317
|||||
1046 TTCTGGGTGAGAGAGAGAGATTCTTCCAGAGATTGACAGCTTCAACCT 1095
317 lGlyTrpProValThrGluAlaGlyHisLeuGlnSerLeuTrpThrGluA 334
|||||
1096 TGGTACCCCGGTACAGGAGCGGGCCACCTGCAAGACCTTACACAGAGA 1145
334 rglauLeuCysHisThrArgAlaLeuProGlyAlaTrpValylsGlyGln 350
|||||
1146 GGCTGCTCTGCATACGCGGGGGGCTGCTGCTGCTGCTGCTGCTGCTGCT 1195

```
351 SerAlaGlyGlyCysArgAsnAsnSerGlyPheProSerAsnProLysPhe 367
|||||
1196 TCAGCAGAGGCTGCCGGAACACAGCGGCTTCCAGCAACCCCAATT 1245
|||||
367 eTrpLeuArgValSerGluProSerGluValTyrIleAlaValLeuGlnA 384
|||||
1246 CTGGCTGGGGTCTCAGAACCGAGTGGGTGATCATTTGGCGTCTGCAGA 1295
|||||
384 rGSerArgLeuHisAlaAlaAspTrpAlaGlyArgAlaArgAlaLeuVal 400
|||||
1296 GATCCAGGCTGCACGGCGGCGACTGGCGACGCGCGCGGCGGCACTGGTG 1345
|||||
401 GlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHisTy 417
|||||
1346 GGTGACAGTCATCTTGTGTGAGCCAGCGAGCATCCCGGCAAGCACTA 1395
|||||
417 rGlnAlaValGlyLeuHisLeuTrpLysValGluLysArgArgValAsnL 434
|||||
1396 CCAGGCTGTGGGTCTGCACCTCTGGAGGTAGAGAGCGGGGTCAATC 1445
|||||
434 euProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAla 450
|||||
1446 TGCTAGGGTCTCTCATGCCCCCGTGGCTGGCACCAGCGTCCCATGCA 1495
|||||
451 TyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTyrTrpLe 467
|||||
1496 TAGCAGCGGAGGTCCACCTGCGTGTGTGCTCTCACCGGCTACTACCT 1545
|||||
467 uAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeuLeuA 484
|||||
1546 GGCTGTCCCCAGCACCTTCTGAGAGCGCGCGGAGTTCCTGCTCC 1595
|||||
484 rGValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAla 500
|||||
1596 GAGTCTTCTCTACCGGGCGAGTCTCCCTTAGCGCCATCAGGCGAGTGCC 1645
|||||
501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyGluTrpGlyThrVa 517
|||||
1646 AGAACACACCCCCGGGCGAGCCCTGCGTGGGGGAGTGGGGACCGT 1695
|||||
517 lGlnLeuArgGlySerTrpArgValGlyGlnThrAlaGlyGlySerArgA 534
|||||
1696 GCAGCTACGGGGTCTTGAGAGAGTCGCCAGACGCGCGGGGAGCAGGGA 1745
|||||
534 snPheAlaSerTyrProThrAsnProCysPheProPheSerValProGlu 550
|||||
1746 ACTTTGCCTCATACCCACCAACCCCTGCTTCCCTTCTCGGTCCCGAG 1795
|||||
551 GlyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArgPr 567
|||||
1796 GGCCCTGGCCCCCGTGGCTCGCATCACTCTGCATCAGCATGCGCGGC 1845
|||||
567 oSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValProG 584
|||||
1846 CAGTGACACCGAGTTCACCCCTCGGCTTCCATATCTTCCAGGTCCCA 1895
|||||
584 luGlyGlyArgSerGlnAspAlaProProLeuLeuGlnGluProLeu 600
|||||
1896 AGGGTGAAGAGCAGGACGACGCCCTCTGCTGCTGCGAGCGCGCTG 1945
|||||
601 LeuSerCysValProHisArgTyrAlaGlnGluValSerArgLeuCysLe 617
|||||
1946 CTGAGCTGGTGGCCACATCGCTACGCCCGCAGGAGGTGAGCGGCTCTGC 1995
|||||
617 uLeuProAlaGlyThrTyrLysValValProSerThrTyrTrpLeuProAsp 634
|||||
1996 CTGCTGCAGGACCTACAGAGTGTGCCCTCCACCTACCTGCCGAGCA 2045
|||||
634 hrGluGlyAlaPheThrValThrIleAlaThrArgIleAspArgProSer 650
|||||
2046 CAGAGGGGGCTTTCAGTGCACCATCGCAACCAAGGATTGACAGGCCATCC 2095
|||||
651 IleHisSerGlnGluMetLeuGlyGlnPheLeuGlnValSerValIle 667
```

```
|||||
2096 ATTCACAGCCAGAGATGCTGGCCAGTTCCTCCAGAGGTCCTCCGTCAT 2145
|||||
667 talAlaValMetLysThr 672
|||||
2146 GGCAGTGATGAAACC 2161
|||||
seq_name: gb_pat:AR153290
seq_documentation_block:
LOCUS AR153290 2620 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 3 from patent US 6235481.
ACCESSION AR153290
VERSION AR153290.1 GI:15120822
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 2620)
AUTHORS Horikawa,Y., Oda,N., Hanis,C.L., Bell,G.I. and Cox,N.J.
TITLE Polynucleotides encoding calpain 10
JOURNAL Patent: US 6235481-A 3 22-MAY-2001;
FEATURES
Location/Qualifiers
source
1..2620
/organism="unknown"
BASE COUNT 448 a 819 c 899 g 454 t
ORIGIN
alignment_scores:
Quality: 3649.00 Length: 672
Ratio: 5.430 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-768-877-2 x AR153290
Align seg 1/1 to: AR153290 from: 1 to: 2620
1 MetArgAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
178 ATGCGGGCGGGCGCGGCGCGAGCGCGGAGGAGCTGTCGGGACGC 227
17 aAlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrProL 34
228 CGCCTTCCCGCGCGGACTCTCGCTCTTCTGCGACTTGTCTACGCGCG 277
34 euAlaGlnPheArgGluAspIleThrTrpArgArgProGlnGluIleCys 50
278 TGGCCCAAGTTCGCGCAGGACATCAGTGGAGCGGCGCCCGCAGAGATTGT 327
51 AlaThrProArgLeuPheProAspAspProArgGluGlyGlnValLysGl 67
328 GCCACACCCCGGCTGTTTCCAGATGACCCAGGGAAGGCGAGTGAAGCA 377
67 nGlyLeuLeuGlyAspCysTyrPheLeuCysAlaCysAlaLeuGlnL 84
378 GGGGCTGCTGGGGGATTGCTGTTCTCTGCTGCTGCTGCGCGCGCTGCA 427
84 ysSerArgHisLeuLeuAspGlnValIleProProGlyGlnProSerTrp 100
428 AGAGCAGCACCTCTCTGGCACCAGGTTCATTCCTCGGGACAGCCGAGCTGG 477
101 AlaAspGlnGluTyrArgGlySerPheThrCysArgIleTrpGlnPheGl 117
478 GCGGACACAGGAGTACCGGGGCTCTCTTCCACTCTCGCATTTGGCAGTTGG 527
117 yArgTrpValGluValThrThrAspAspArgLeuProCysLeuAlaGlyA 134
528 ACGTGGGTGGAGGTGACACAGATGACCCGCTGCGCTGCGCTTGCAGGGA 577
134 rGLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
|||||
```

578 GACTCTGTTTCCCGCTCCAGAGGAGATGTTCTGGCTCCCTTA 627
 151 LeuGluLysValTyrAlaLysValHisGlySerTyrGluHisLeuTyrAla 167
 628 CTGCAAAAGGCTTACGCCAGGTCATGGTCTTACGAGCACCTGTGGGC 677
 167 aglyGluValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGlu 184
 678 CGGGAGGTGGCGGATGGCTGGTGGACTGACCGCGGCTGGCAGAAA 727
 184 rgtPheSerLeuLysGlyValAlaGlySerGlyGlyGlnAspArgPro 200
 728 GATGGAACTTCAAGGGCTTACAGGAAAGCGAGCCACAGGACAGGCCA 777
 201 GlyArgTyrPheGluHisArgThrCysArgGlnLeuHisLeuLysAspGlu 217
 778 GGGCGCTGGAGCACAGGACTTGTGGGAGCTGCTCCACCTGAAGACCA 827
 217 nCysLeuLysSerCysCysValLeuSerProArgAlaGlyAlaArgGln 234
 828 GTGTCTGATCAGCTGCTGCTGCTCAGCCAGAGCAGGTGCCGGAGC 877
 234 euGlyGluPheHisAlaPheIleValSerAspLeuArgGlnLeuGlnGly 250
 878 TGGGGAGATTCACATGCTTCAATGCTCTGACCTGCGGAGCTCCAGGGT 927
 251 GlnAlaGlyGlnCysIleLeuLeuLeuArgIleGlnAsnProTyrPheArg 267
 928 CAGCGGGGCCAGTGCATCTCTGCTGGATCCAGAAACCTTGGGGCG 977
 267 gArgCysTyrPheGlnGlyLeuTyrArgGlnGlyGlyGlnGlyTyrSerGln 284
 978 GCGGTGCTGGCAGGGGCTCTGAGAGAGGGGGTGAAGGTGAGGAGCCAG 1027
 284 alaPheAlaAlaValAlaSerGluLeuLeuSerGlnLeuGlnGlnGlyGlu 300
 1028 TAGATGACAGCGGTACATCTGATCTCTCCAGCTCCAGGAAGGGGAG 1077
 301 PheTyrValGlnGlnGluGluPheLeuArgGlnPheAspGluLeuThrVal 317
 1078 TTCTGGGTGGAGAGAGAGACTTCTCAGGAGTTTACGACCTCACCGT 1127
 317 IGIYTrProValThrGluAlaGlnHisLeuGlnSerLeuTyrThrGlu 334
 1128 TGGCTACCCGGTACAGGAGCGGCGCCACCTCAGAGCTCTACACAGGA 1177
 334 rGluLeuLysHisThrArgAlaLeuProGlyAlaTyrValGlyGln 350
 1178 GCGTCTCTGCAATACGGGGCTGCTGGGCTGGGTCAAGGGCCAG 1227
 351 SerAlaGlyGlyCysArgAsnAsnSerGlyPheProSerAsnProLysPhe 367
 1228 TCAGCAGAGGCTCCCGAACAACAGCGGCTTTCCAGCAACCCCAATT 1277
 367 eTyrLeuArgValSerGluProSerGluValTyrIleAlaValLeuGln 384
 1278 CTGGCTCGGGCTCAGAACCGAGTGAAGGTATGAGCGCTCTGCGAGA 1327
 384 rGSerArgLeuHisAlaAlaAspTyrAlaGlyArgAlaValLeuVal 400
 1328 GATCCAGGCTGACAGCGCGGCTGAGGCGGCGGCGGCGGCTGAGTG 1377
 401 GlysPheSerHisThrSerTyrPheProAlaSerIleProGlyLysHisTyr 417
 1378 GGTACAGCTATCTGCTGTGAGGCCACGACGATCCCGGCGCAAGCCTA 1427
 417 rGlnAlaValGlyLeuHisLeuTyrPheValGluLysArgValAsnL 434
 1428 CCAAGGCTGGGTCTGACCTCTGAAGGTGAAGAGCGGGGCTCAATC 1477
 434 euProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAla 450
 1478 TGGCTAGGCTGCTGTCATGCCCCCGTGGCTGGCACCGGGTGGCCATGCA 1527

451 TyrAspArgGluValHisLeuArgCysGlnLeuSerProGlyTyrTyrLe 467
 1528 TACAGCCGGAGGCTCCACCTTCTGTGAGCTTCCACCGGGCTACTACT 1577
 467 uAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeu 484
 1578 GGCTGTCCCGACACCTTCTGAAGAGCGCGCAGGGAGTCTGCTCC 1627
 484 rValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAla 500
 1628 GAGCTTCTTACCGGGGAGCTTCCCTTACGGCATAGGCGAGTGCC 1677
 501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyIuTyrPheVal 517
 1678 AAGAACACACCCCGGGGAGCCCTGCTGGGGGAGTGGGGACCGT 1727
 517 IGIleuArgGlySerTyrPheValGlyGlnThrAlaGlyLysSerArg 534
 1728 GCACCTACGGGCTTCTTGAAGATCGCCAGACGGGGGGGACAGGA 1777
 534 snPheAlaSerTyrProThrAsnProCysPheProPheSerValProGlu 550
 1778 ACTTGGCTCATACCCCAACACCCCTGCTTCCCTTCTGCTCCGAG 1827
 551 GlyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArgPro 567
 1828 GGGCTGGCCCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1877
 567 oSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValProG 584
 1878 CAGTGACACCGAGTTCACACCCATCGGCTTCATATCTTCCAGSTCCAG 1927
 584 IuGlyGlyArgSerGlnAspAlaProProLeuLeuLeuGlnGluProLeu 600
 1928 AGGCTGAGAGGACGACGAGGACGACCCGCTGCTGCTGCTGCTGCTGCT 1977
 601 LeuSerCysValProHisArgTyrAlaGlnGluValSerArgLeuCysLe 617
 1978 CTGAGCTGCGTCCACATCGCTACGCGCAGAGGTGAGCGGCTCTGCT 2027
 617 uLeuProAlaGlyThrTyrLysValValProSerThrTyrLeuProAsp 634
 2028 CCTCGTGCAGGACCTACAAAGTTTGGCTTCCACCTACTGCGGAGCA 2077
 634 hrGluGlyAlaPheThrValThrIleAlaThrArgIleAspArgProSer 650
 2078 CAGAGGGGGCTTACAGTGCATGCCAACCAAGATGAGACAGCCATCC 2127
 651 IleHisSerGlnGluMetLeuGlnPheLeuGlnGluValSerValMe 667
 2128 ATTCACAGCCAGGAGATGCTGGGCACTTCTCCAAAGGTCTCGTCAT 2177
 667 tAlaValMetLysThr 672
 2178 GGCAGTATGATAAAC 2193

seq_name: gb_pr:AF089088

seq_documentation_block:

LOCUS AF089088 2620 bp mRNA linear PRI 23-JAN-2001

DEFINITION Homo sapiens calpain-like protease CAPN10a mRNA, complete cds.

ACCESSION AF089088

VERSION AF089088.1 GI:10503934

KEYWORDS

SOURCE

ORGANISM

human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 2620)
 Horikawa,Y., Oda,N., Cox,N.J., Li,X., Orho-Melander,M., Hara,M.,
 Hinokio,Y., Lindner,T.H., Mashima,H., Schwarz,P.E., del
 Bosque-Plata,L., Horikawa,Y., Oda,Y., Yoshuchi,I., Collina,S.,

Polonsky, K.S., Wei, S., Concannon, P., Iwasaki, N., Schulze, J.,
Baier, L.J., Bogardus, C., Groop, L., Boerwinkle, E., Hanis, C.L. and
Bell, G.I.
Genetic variation in the gene encoding calpain-10 is associated
with type 2 diabetes mellitus
Nat. Genet. 26 (2), 163-175 (2000)
20472315
2 (bases 1 to 2620)
Horikawa, Y. and Bell, G.I.
Direct Submission
Submitted (02-SEP-1998) Howard Hughes Medical Institute, The
University of Chicago, 5841 S. Maryland Avenue, MC1028, Chicago, IL
60637, USA

FEATURES

source
1..2620
/organism="Homo sapiens"
/db.xref="taxon:9606"
178..2196
/note="type A"
/codon_start=1
/product="calpain-like protease CAPN10a"
/protein_id="AAG17966.1"
/db.xref="GI:10503935"
/translation="MRAGGATPARELFRDAFPADSSLFCDLSTPLAQREDITWR
RPEICATPRFPDDREGQVKGLGDCWFLCACAALQKSRHLLDQVIPGQPSWAD
QYRGSTFCRIWQGRWVEVTDRLPLAGRLCFSRQREDVFWLPLERKYAKVHG
SYELHAGQVADALVDLTGGLAERNLKGAGSGQDQGRWEHRTCRQLHLKDDC
LISCCVSPRAGARELGEFHAFIVSDRLRELQAGQCTILLRIQNPGRRCWQGLWRE
GEGMSQVAAVASSELLSQEGEFVEEFELREFDELTVGYVPTENGHLQSLYTER
LLCHTRALPCAWKVGSGAGGRNNSGPPNPKFWLRVSESEYVIAVLRSLRHAADW
AGRALVGDSSHTSNFASIPGKHVQAVLHLWKERRVNLPRVLSMPVPVAGTACHA
YDREHLRCELSPGYLVAVPTFLKDPAGFELRVSTGRVSLSAIRAVAKNTTPGAA
LPAGEWTVQLRGWSRGVQTAGGSNFASFYTNPCFPFVPEGPGRCVRITLHQHCR
PSDEFHPHIGFIQVPEGGRSQDAPLLQLQELLSVPHRYAOEVSRICLLPAGTVK
VVPSTYLPDTEGAFVTTIATRDRPSIHSQELQFLQEVSMVAMKT"

polyA_signal
BASE COUNT 448 a 819 c 899 g 454 t
ORIGIN

alignment_scores

Quality: 3649, 00 Length: 672
Ratio: 5, 430 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block

US-09-768-877-2 x AF089088

Align seg 1/1 to: AF089088 from: 1 to: 2620

1 MetArgAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
|||||
178 ATGCGGGCGGGCGGGCGGCGAGCGCGGAGGAGCTGTCGGGAGCGC 227

17 aAlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrProL 34
|||||
228 CGCTTCCCGCGCGGAGACTCTCGCTCTTCGCGACTTGTCTACGCGCG 277

34 euAlaGlnPheArgGluAspIleThrTrpArgProGlnGluIleCys 50
|||||
278 TGGCCCAAGTTCGCGAGGACATCAGTCGAGGCGGCGCCCGAGGATTGCT 327

51 AlaThrProArgLeuPheProAspAspProArgGluGlnValIysG1 67
|||||
328 GGCACACCCCGCGGTGTTCCAGATGACCCAGCGGAGGCGAGGTGAAGCA 377

67 nGlyLeuLeuGlyAspCysTrpPheLeuCysAlaCysAlaAlaLeuGlnL 84
|||||
378 GGGGCTGCTGGGGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 427

84 ySerArgHisLeuLeuAspGlnValIleProProGlyGlnProSerTrp 100
|||||
428 AGAGCAGGCACCTCTCTGGACACAGTCACTTCTCCGGGACAGCGGAGCTGG 477

101 AlaAspGlnGluTrpArgGlySerPheThrCysArgIleTrpGlnPheG1 117
|||||
478 GCCGACGAGAGTACCGGGCTCTTACCTGTCGATTTGGCAGTTGG 527

117 YArgTrpValGluValThrThrAspAspArgLeuProCysLeuAlaGlyA 134
|||||
528 ACCTGGTGGAGGTGACACAGATGACCGCTGCGCTGCTTGCAGGGA 577

134 rGLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
|||||
578 GACTCTGTTCTCCCGCTGCCAGAGGAGGATGTTCTGGCTCCCTTA 627

151 LeuGluLysValTrpAlaLysValHisGlySerTyrGluHisLeuTrpAl 167
|||||
628 CTGGAAGAAGTCTACGCCAAGTCCATGGTCTCTACGACACCTGTGGGC 677

167 aGlyGlnValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGluA 184
|||||
678 CGGCGAGTGGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 727

184 rGTrpAsnLeuLysGlyValAlaGlySerGlyGlnGlnAspArgPro 200
|||||
728 GATGGAACCTGAAGGCGGTAGCAGGAAGCGGAGCGGAGGACAGGCGCA 777

201 GlyArgTrpGluHisArgThrCysArgGlnLeuLeuHisLeuLysAspG1 217
|||||
778 GCGCGTGGGACACAGGACTTGTGCGCAGCTCTCCACTGAAGGACCA 827

217 nCysLeuIleSerCysCysValLeuSerProArgAlaGlyAlaArgGluL 234
|||||
828 GTGCTCTGATCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 877

234 euGlyCyluPheHisAlaPheIleValSerAspLeuArgGluLeuGlnGly 250
|||||
878 TGGGGGAGTTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 927

251 GlnAlaGlyGlnCysIleLeuLeuLeuArgIleGlnAsnProTrpGlyAr 267
|||||
928 CAGCGGGCGGAGTGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 977

267 gArgCysTrpGlnGlyLeuTrpArgGluGlyGlyGlyGlyGlyTrpSerGlnV 284
|||||
978 GCGGTGCTGGCAGGGGCTCTGGAGAGAGGGGGGTGAAGGCTGAGGCGAG 1027

284 aAlaAspAlaValAlaSerGluLeuLeuSerGlnLeuGlnGluGlyGlu 300
|||||
1028 TAGATGACCGGTAGCATCTGAGCTCTCTGCTGCTGCTGCTGCTGCTGCTG 1077

301 PheTrpValGluGluGluPheLeuArgGluPheAspCyluLeuThrVa 317
|||||
1078 TTTCTGGTGGAGGAGGAGGAGTTCCTCAGGAGTTTGACGAGCTCACCGT 1127

317 lGlyTrpProValThrGluAlaGlyHisLeuGlnSerLeuTrpThrGluA 334
|||||
1128 TGGCTACCGCGGTACGAGGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1177

334 rGLeuLeuCysHisThrArgAlaLeuProGlyValaTrpValLysGlyGln 350
|||||
1178 GGCTGCTCTGCCATACGCGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1227

351 SerAlaGlyGlyCysArgAsnAsnSerGlyPheProSerAsnProLysPh 367
|||||
1228 TCAGCAGGAGGCGTGCAGGAAACAGCGGCTTTTCCAGCAACCCCAATT 1277

367 eTrpLeuArgValSerGluProSerGluValTrpIleAlaValLeuGlnA 384
|||||
1278 CTGGCTGCGGGTCTCAGAACCGAGTGAGGTGTACATTGCGCTCTCTCAGA 1327

384 rGSerArgLeuHisAlaAlaAspTrpAlaGlyArgAlaArgAlaLeuVal 400
|||||
1328 GATCCAGGCTGCACCGCGGAGTGGGAGGCGCGGCGGCGGCGGCGGCGG 1377

```

401 GlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHisIty 417
1378 GGTGACAGTACTCTCGTGGAGCCAGGACCATCCGGGACACAGCA 1427
417 rGlnAlaValIglyLeuHisLeuTrpLysValIcLulysArgArgValAsnL 434
1428 CCAGGCTGGGGTGTGCACCTCTGAGAGGTAGAGAAAGCGCGGCTCATATC 1477
434 euProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAla 450
1478 TGCCCTAGGGTCTCTGTCATATGCCCGGTGGCTGGCAAGCGGTCATGCA 1527
451 TyrAspArgIuValHisLeuArgCysGluLeuSerProGlyTyrTyrLe 467
1528 TAGACCGGGAGGTCCACCTCGTGTGAGCTCTCACCGGGCTACTACTCT 1577
467 uAlaValProSerThrPheLeuLysAspAlaProGlyLupheLeuAla 484
1578 GGCTGTCCCGACACCTTCTCTGAGAGGAGCGCCAGGGAGTTCCTGCTCC 1627
484 rValPheSerThrGlyValArgValSerLeuSerAlaIleArgAlaValAla 500
1628 GAGTCTCTCTACGGGGGAGGTCTCCCTTACGCGCATCCAGGAGAGTGGCC 1677
501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyLupTrpGlyThrVa 517
1678 AAGAAGACACACCCCGGGGAGACCCCTGCTGGGGGAGTGGGGAGCCGT 1727
517 IGlLeuArgGlySerTrpArgValIglyLlnThrAlaGlyLysArga 534
1728 GCAGCTAGGGGTTCTGTGAGAGTGGCCAGACGGGGGGGAGACAGAGA 1777
534 snPheAlaSerTyrProThrAsnProCysPheProPheSerValProGlu 550
1778 ACTTGGCTCATACCCACCAACCCCTGCTTCCCTTCTGGTCCCGAG 1827
551 GLyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArgPr 567
1828 GGCCCTGGCCCCCGCTGGCTCCGATCCTGATCAGCACTGGCGGCC 1877
567 oSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValProG 584
1878 CAGTGACACCGAGTTCACCCCATCGCTTCATATCTTCCAGGCTCCAG 1927
584 IuGlyGlyArgSerGlnAspAlaProProLeuLeuGlnGluProLeu 600
1928 AGGGTGGAGAGAGCCAGACACCCACCTGCTGCTCAGAGCGGCTG 1977
601 LeuSerCysValProHisArgTyrArgIleGlnGluValSerArgLeuCysLe 617
1978 CTGAGCTCGTGGCCATCGCTAGCGCCAGAGGTGAGCGGCTGTGCT 2027
617 uLeuProAlaGlyThrTyrLysValValProSerThrTyrLeuProAspT 634
2028 CCTGCTGTCAGACCTACAGAGTGTGCTCTCCACCTACCTGGCGGACA 2077
634 hrGluGlyAlaPheThrValThrIleAlaThrArgIleAspArgProSer 650
2078 CAGAGGGGGCTTCCACAGTGAACATCGCAACCCAGATTGACAGGCAATCC 2127
651 IleHisSerGlnGluMetLeuGlyGlnPheLeuGlnGluValSerValMe 667
2128 ATTCAACAGCCAGAGATCTGGGCCAGTTCCTCCAAAGAGTCTCCGTAT 2177
667 tAlaValMetLysThr 672
2178 GGCAGTGAATAAAC 2193
seq_name: gb_pat:ARI53295
seq_documentation_block: 2455 bp DNA linear PAT 08-AUG-2001
LOCUS ARI53295
DEFINITION Sequence 13 from patent US 6235481.

```

```

ACCESSION ARI53295
VERSION ARI53295.1 GI:15120827
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2455)
AUTHORS Horikawa, Y., Oda, N., Hanis, C. L., Bell, G. I. and Cox, N. J.
TITLE Polynucleotides encoding calpain 10
JOURNAL Patent: US 6235481-A 13 22-MAY-2001;
FEATURES
source 1..2455
BASE COUNT 429 a 769 c 827 g 430 t
ORIGIN

alignment_scores:
Quality: 3604.00 Length: 673
Ratio: 5.395 Gaps: 2
Percent Similarity: 99.257 Percent Identity: 99.257

alignment_block:
US-09-768-877-2 x ARI53295
Align seg 1/1 to: ARI53295 from: 1 to: 2455

1 MetArgAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
42 ATGCGGGGGGGCGGGGGCGAGCGCGGGGAGAGAGTGTTCGGGAGCGC 91
17 aAlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrPro 34
92 GCCTTCCCGCGCGGAGCTCTGCTCTTGTGCACTTGTGTACGCCGC 141
34 euAlaGlnPheArgGluAspIleThrTrpArgArgProGlnGluIleCys 50
142 TGCCCGAGTCCGAGAGCATCAGTGGAGCGGCCCGCCAGAGATTTC 191
51 AlaThrProAlaGluPheProAspAspProArgGluGlyGlnValLysG 67
192 GCCACACCCCGGCTGTTCCAGATGACCCAGGAGGAGGAGTGAAGCA 241
67 nGlyLeuGluGlyAspCysTrpPheLeuCysAlaCysAlaAlaLeuGln 84
242 GGGGCTGGGGGATTCCTGTTCTGTGCTCTGCGCGCGGCTGCAGA 291
84 ySerArgHisLeuLeuAspGlnValIleProProGlyGlnProSerTrp 100
292 AGAGAGAGCACCCTCTGAGACCAAGTCAATTCCTCCGGACAGCGAGCTGG 341
101 AlaAspGlnGluTyrArgIlySerPheThrCysArgIleTrpGlnPheG 117
342 GCCGACACGAGAGTACCGGGGCTCTTCACTGCTGCAATTTGGCAATTTGG 391
117 YArgTrpValGluValIleThrAspAspArgLeuProCysLeuAlaGly 134
392 ACGTGGGTGGAGTGACACAGATGACCGCTGCGCTTGCCTTGCAGGGA 441
134 rGluCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
442 GACTGTGTTTCTCCGCTGCCAGAGGAGAGATGTTCCTGCTCCCTTA 491
151 LeuGluLysValTyrAlaLysValHisGlySerTyrGlnHisLeuTrpA 167
492 CTGGAA.....AAGGTTCATGGGTCTTACAGACACCTGTGGG 529
167 IaglyGlnValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGlu 183
530 CCGGCGAGGTGGCGGATGCCCTGTGGACCTACACGGCGGCTGGCAGAA 579
184 ArgTrpAsnLeuLysGlyValAlaGlySerGlyGlyGlnGlnAspArgPr 200

```

580 AGATGGAACTCTGAAGGGCTAGCAGGAACGGAGGCCACGAGGACAGGCC 629

200 oGlyArgTtPgluHisArgThrCysArgGlnLeuLeuHisLeuLysaspG 217

630 AGCCCGCTGGGAGCACAGACATTGTCTGCAGCTGCTCCACCTGAAGGACC 679

217 InCysLeuLeuSerCysCysValLeuSerProArgAlaGlyAlaArgGlu 233

680 AGTGTCTGATCAGCTGCTGCTGCTCAGCCACGACAGAGTGCCTGGGAG 729

234 LeuGlyGluPheHisAlaPheIleValSerAspLeuArgGluLeuGlnGln 250

730 CTGGGGAGGATTCATATGCCCTTCATTCTCTCGGACCTCGCGAGCTCCAGG 779

250 yGlnAlaGlyGlnCysIleLeuLeuLeuArgIleGlnAsnProTrpGlyA 267

780 TCAGCGGGCCAGTAGTCATCTGCTCTCGGATCCAGAACCCCTGGGGGCC 829

267 rGArgCysTrpGlnGlyLeuTrpArgGluGlyGlyGlyGlyTrpSerGln 288

830 GCGGCTGCTGCGAGGGCTCTGACAGAGGGGGTGAAGGCTGGAGCCAG 879

284 ValAspAlaAlaValAlaSerGluLeuLeuSerGlnLeuGlnGluGlyGln 300

880 GTAGATGTCACGGGTAGCATCTGAGCTCTGTCCTCCAGCTTCCAGGAAGGGA 929

300 uPheTrpValGluGluGluPheLeuArgGluPheAspGlnLeuThrV 317

930 GTTCTGGGTGGAGAGAGGAGATTCTCTCAGGGAGTTTGACGAGCTCAACG 979

317 aGlyTyrProValThrGluAlaGlyHisLeuGlnSerLeuTyrThrGlu 333

980 TTGGGTACCCGGTCTCAGAGGCGGGCCCTGCAGAGCCTCTACACAGAG 1029

334 ArgLeuLeuCysHisThrArgAlaLeuProGlyAlaTrpValLysGlyGln 350

1030 AGGCTGTCTTGCCATACCGGGGGCTGCTCTGGGGCTTGGTCAAGGGCCA 1079

350, nSerAlaGlyGlyCysArgAsnAsnSerGlyPheProSerAsnProLysP 367

1080 GTACAGAGAGGCTCGCCGGACACAGCGGGCTTTCCAGCAACCCCAAAAT 1129

367 heTrpLeuArgValSerGluProSerGluValTyrIleAlaValLeuGln 383

1130 TCTGGCTCGGGTCTCAGAACCGAGTGAGGTGATCGCTGCTCTCAG 1179

384 ArgSerArgLeuHisAlaAlaAspTrpAlaGlyArgAlaArgAlaLeuVa 400

1180 AGATCCAGCTGTCACGGCGGAGCTGGGACGCGCGCGCGCGGCGCACGTGT 1229

400 lGlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHisT 417

1230 GGGTGACAGTCATCTCTGCTGGAGCCCGAGGATCCCGGGCAAGGACCT 1279

417 yTGlnAlaValGlyLeuHisLeuTrpLysValGluLysArgArgValAsn 433

1280 ACCAGGCTGTGGGTCTGCACCTCTGGAAGGTAGAGAAGCGCGGGGTCAAT 1329

434 LeuProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAl 450

1330 CTGGCTCAGGGTCTGTCCATGCCCGCGTGGCTGGCACCGCGCTGCCATGC 1379

450 aTyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTyrTyrL 467

1380 ATACCAACGGGAGGTCCACCTTCGCTTGTGAGCTCTCACCGGGCTACTACC 1429

467 euAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeuLeu 483

1430 TGGCTGTCTCCAGCACCTTCTGAAGGACGCGCGAGGGAGTTCCTGCTC 1479

484 ArgValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAl 500

1480 CAGATGTTCTCTACCGGGCGAGTCTCCTTTAGCGCCATCAGGCGATGGC 1529

500	alysAsnThrThrProGlyAlaAlaLeuProAlaGlyGluTrpGlyThrV	517
1530	CAAGAAACACCAACCCCGGGGAGCCCTGCCTCGCGGGAGTGGGGACCG	1579
517	alGlnLeuArgGlySerTrpArgValGlyGlnThrAlaGlyGlySerArg	533
1580	TGCAGCTACGGGGGTCCTGGAGAGTCGGCCAGACGGCGGGGGCAGCAGG	1629
534	AsnPheAlaSerTyrProThrAsnProCysPheProPheSerValProGln	550
1630	AACTTTGCTCATACCCACCAACCCCTGCTTCCCTTCCTCGGTCCCGA	1679
550	uGlyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArgP	567
1680	GGGGCCTTGGCCCCCGCTGCGTCGCGATCACTCTGCAATCAACACTGCCGC	1729
567	roSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValPro	583
1730	CCAGTGACACGGAGTTCCACCCCATCGGCTTCCATATCTTCCAGGTCCCA	1779
584	GluGlyGlyArgSerGlnAspAlaProProLeuLeuLeuGlnGluProLe	600
1780	GAGGCTGGAAGAGCCACGAGACGACCCCCACTGCTGCTGAGGAGCGCT	1829
600	uLeuSerCysValProHisArgTyrAlaGlnGlnValSerArgLeuCysL	617
1830	GCTGAGCTGGGTGCCATCGCTACGCCCCAGGAGGTGAGCCGCTCTGCC	1879
617	euLeuProAlaGlyThrTyrLysValValProSerThrTyrLeuProAsp	633
1880	TCCTGCCTGCAGGCACCTACAAGGTTGTGCCCTCCACCTACCTGCCGGAC	1929
634	ThrIuGlyAlaPheThrValThrIleAlaThrArgIleAspArgProse	650
1930	ACAGGGGGCGCTTCACAGTGCACATCGCAACCGAGATTGACAGGCATC	1979
650	rIleHisSerGlnGluMetLeuGlyGlnPheLeuGlnGluValSerValM	667
1980	CATTACAGCCAGGAGATGCTGGGCCAGTTCTCTCCAGAGGCTCTCGTCA	2029
667	etalValMetLysThr	672
2030	TGGCAGTGATGAAAAACC	2046

seq_name: qb_pr:AF089094

seq_documentation_block:

LOCUS	AF089094	2455 bp	mRNA	linear	PRI 23-JAN-2001
DEFINITION	Homo sapiens calpain-like protease CAPN10f mRNA, complete cds.				

ACCESSION	AF089094
VERSION	AF089094.1

KEYWORDS

SOURCE human

ORGANISM Homo :

Eukar:

Mamma.

REFERENCE

AUTHORS

Hinok.

Bosqu

Polon:

batter

TITLE	Cont.	bell, (

THE GELLEC with

JOURNAL. Nat.

JOURNAL OF MEDICAL
MEDICINE
20472

REFERENCE

AUTHORS Horioka

[illegible]JOURNAL
Submit

Univer

60637, USA

FEATURES

Source

Location/Qualifiers
1..2455
/organism="Homo sapiens"
/db_xref="taxon:9606"

CDS

42..866
/note="type F; alternatively spliced"
/codon_start=1
/product="calpain-like protease CAPN10f"
/protein_id="A017972.1"
/db_xref="GI:10503948"
/translation="MRAGKATPARELFNDAPPAADSLFCDLSTPLAQFREDITWR
RPEICATPPLPDDPREQVKGILGDCWFLCACAALQKSRHLLDQVLPQGPSPVAD
OERYSGFTCRIMQFRMVEYTDRLPLCLAGRCFSRCQREDVFWLPLEKGFVTLRA
PVGRAGGCPGCPDRRGRKMEPEGSRKRRPAGCARPLGAODLSAAPPDEGVSDOL
LRAPQSRCPGAGVPCILHCLCPAGAGSGGVPVHPAADPEPLGPAVLALERGC"

BASE COUNT 429 a 769 c 827 g 430 t
ORIGIN

alignment_scores:

Quality: 3604.00 Length: 673
Ratio: 5.395 Gaps: 2
Percent Similarity: 99.257 Percent Identity: 99.257

alignment_block:

US-09-768-877-2 x AF089094

Align seg 1/1 to: AF089094 from: 1 to: 2455

```

1 MetcrgalaglyargglyalathProalaargluleupheargaspal 17
|||||
42 ATGCGCGCGGCGCGGCGCGACCGCGCGAGAGAGCTGCCGGGAGCG 91
|||||
17 aAlaPheProAlaAlaAspSerSerleupheCysAspLeuSerThrProL 34
|||||
92 CGCCTTCCCGCGCGCGGACTCTGCTCTGCGACTTGTCTACGCGCGC 141
|||||
34 euaAlaGlnPhearggluaspllethrTrpArgArgProGlnGluileCys 50
|||||
142 TGGCCAGTTCGCGAGGACATCACTGAGAGCGGCCCGCCAGAGATTGT 191
|||||
51 AlaThrProArgLeupheProAspAspProArgGlnGlnValLysGlu 67
|||||
192 GCCACACCGCGCGTGTTCAGATGACCCAGGAGGAGGAGTGAAGCA 241
|||||
67 nglyLeuLeuGlnLysPcysTrpPheLeuGlnLysAlaAlaLeuGlnL 84
|||||
242 GGGGCTCTGGGGGATTGCTGTGCTGTGCTGTGCGCGCGCTGCAGCA 291
|||||
84 ySSerArgHISleuLeuAspGlnValIleProProGlnProSerTrp 100
|||||
292 AGAGCAGGACACTCTGACACAGGTCACTTCTCCGGGACAGCCAGGCTGG 341
|||||
101 AlaAspGlnGluTrpArgLysPheThrCysArgLleThrGlnPheGlu 117
|||||
342 GCCAGCAGAGGATACCGGGGCTCTTCACTGCGCATTTGGCAGTTGG 391
|||||
117 yArgTrpValGluValIleThrAspAspArgLeuProCysLeuAlaGlu 134
|||||
392 ACCTGTGGTGAGGTGACACAGATGACCGCTCTCCGCTTCCAGGGA 441
|||||
134 rgleuCysPheSerArgCysGlnArgLnuAspValPheThrLeuProLeu 150
|||||
442 GACTCTTTTCTCCCTGCGCAGAGGAGATGTTCTGCTCCCTTA 491
|||||
151 leuGlnLysValTyralalysValHisGlySerTyrGlnHisLeuTrp 167
|||||
492 CTGGAA.....AAGGGTCATGGGTCTTACGAGACCTGGGG 529
|||||
167 laGlnGlnValAlaAspAlaLeuValAspLeuThrGlnGlyLeuAlaGlu 183
|||||
530 CGGGGAGGTGGGGATGCCCTGTGAGACCTGACCGGCGCTGGCAGAA 579

```

```

184 ArgTrpAsnLeuLysGlyValAlaGlySerGlyGlnGlnAspArgPr 200
|||||
580 AGATGGACCTGAAAGCGCTAGCAGAAAGCGGACCGACGACAGCC 629
|||||
200 oGlyArgTrpGlnHisArgTrpCysArgGlnLeuLeuHisLeuLysAsp 217
|||||
630 AGGCGGCTGGAGCAGAGACTTGTGGCAGACTCTCCACTGAAGGACC 679
|||||
217 lncysleuIleSerCysCysValleuSerProArgAlaGlyAlaArgGlu 233
|||||
680 AGTCTGTATCAGCTGCTGCTGCTCAGACCCAGAGCAGCTGCCGGGAG 729
|||||
234 leuGlyGlnPheHisAlaPheIleValSerAspLeuArgGlnGlnGlu 250
|||||
730 CTGGGGGAGTTCCATGCTTCAATTGTCTCGGACCTGGGAGCTCCAGG 779
|||||
250 yGlnAlaGlylncysIleLeuLeuLeuArgIleGlnAsnProTrpGly 267
|||||
780 TCAGGCGGCGCAGTGCATCTGCTGCTGGGATCCAGAACCCCGGGGCC 829
|||||
267 rArgGlyTrpGlnGlyLeuThrArgGlnGlyGlyGlyTyrPheSerGln 283
|||||
830 GCGGCTGTGGCAGGGGCTCTGAGAGAGGGGGGTGAAGGCTGGAGCCAG 879
|||||
284 ValAspAlaAlaValAlaSerGlnLeuLeuSerGlnLeuGlnGlnGly 300
|||||
880 GTAGATGACGCGGTAGCATCTGACTCTCTGCTCCAGCTCCAGAAAGGGA 929
|||||
300 uPheTrpValGlnGlnGlnGlnPheLeuArgGlnPheAspGlnLeuThr 317
|||||
930 GTTCTGGGTGAGAGAGGAGGATTCCTCAGAGGATTTGACAGCTCAACG 979
|||||
317 aGlyTyrProValThrGlnAlaGlyHisLeuGlnSerLeuTyrThrGlu 333
|||||
980 TTGGCTACCGCGTCCAGGAGCGCGCCACCTGCGAGAGCTCTTACACAG 1029
|||||
334 ArgLeuLeuCysHisThrArgAlaLeuProGlnValAlaTrpValLys 350
|||||
1030 AGCTGCTCTCTCCATACCGCGGCGCTGCTGGGCTGGGTCAGAGGCCA 1079
|||||
350 nSerAlaGlyGlyCysArgAsnAsnSerGlyPheProSerAsnProLys 367
|||||
1080 GTCAGCAGAGAGGTCGCGGAACAACAGCGGCTTCCAGACACCCCAAT 1129
|||||
367 heTrpLeuArgValSerGluProSerGlnValTyrIleAlaValLeuGln 383
|||||
1130 TCTGCTGCGCGGCTCAGAACCGAGTGAAGGTACATTTGCCGCTCCAG 1179
|||||
384 ArgSerArgLeuHisAlaAlaAspTrpAlaGlyArgAlaArgAlaLeu 400
|||||
1180 AGATCCAGGCTGCACGCGGCGACTGGGACGCGCGCGGACATCGT 1229
|||||
400 lGlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHis 417
|||||
1230 GGGGACAGTATCATCTCTGTGAGACCCAGCAGATGCCGGGCAAGCACT 1279
|||||
417 yGlnAlaValGlyLeuHisLeuThrLysValGlnLysArgArgValAsn 433
|||||
1280 ACCAGGCTGTGGGTCTGACCTGTGAAGGTAGGAAGCGGGGTCAT 1329
|||||
434 leuProArgValleuSerMetProProValAlaGlyThrAlaCysHisAl 450
|||||
1330 CTGCTTGAAGGCTCTGCTCATGCCCCCGCTGGCTGGACCCGCTCCATGC 1379
|||||
450 aTyrAspArgGlnValHisLeuArgCysGlnLeuSerProGlyTyrTyrL 467
|||||
1380 ATACGACCGGAGGTCACCTGCTTGAAGCTTCAACCGGGCTACATACC 1429
|||||
467 euaAlaValProSerThrPheLeuLysAspAlaProGlyGlnPheLeuLeu 483
|||||
1430 TGGCTGTCCACAGACCTCTCTGAAGAGCGGCGGAGGATTCCTGCTC 1479

```



```

337 GCCATCGCCCGGTTGTTCCAGATGACCCGACAGGAGGCGAGTGAACA 386
67 nglYleuLeuGlyAspCysTrpPheLeuCysAlaCysAlaAlaLeuGln 84
387 GGGGCGCTGGGGGATTCGTGTTCTGTGTGCTGTGGCCGCTGCACA 436
84 ySerArghisLeuLeuAspGlnValIleProProGlyInProSerTrp 100
437 AGAGCAGGACCTCTGTGAACAGATGATCTCCAGAGACGCCAAGCTGG 486
101 AlaAspGlnGluTrpArgGlySerPheTrpCysArgIleTrpGlnPheG 117
487 GCCGACAGAGTACCAGGCGCTCTTCACCTGCCATTGGCAGATTGG 536
117 yArgTrpValGluValThrThrAspAspArgLeuProCysLeuAlaGly 134
537 AGCTGGGTGGAGGTGACCAAGATGACCGCTGCCATGCTGCAGGGA 586
134 rglLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
587 GACTCGTTTCCCGCTGCCAGAGGAGATGTGTCTGGCTCCCTTA 636
151 LeuGluLysValTyrAlaLysValHisGlySerTyrGluHisLeuTrpAl 167
637 CTGGAAAGCTACGCGCCAGGCTCATGGGTCTACGAGACCTGTGGGC 686
167 aglYglnValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGln 184
687 AGGCAAGGTGGCGGATGCTCTGTGTGACTGTGCTGCTGCAGGAAA 736
184 rgtPAsnLeuLysGlyValAlaGlySerGlyGlyGlnAspArgPro 200
737 GATGAGCCTGAAAGGCGGTAGCAGAGAGCGGAGCGCAGACAGCGCTG 786
201 GlyArgTrpGlnHisArgThrCysArgGlnLeuLeuHisLeuLysAspG 217
787 GCGCCGTGGAGACAGACGCTGTGGCAGCTCCCGCTGMAAGACCA 836
217 nCysLeuIleSerCysCysValLeuSerProArgAlaGlyAlaArgGln 234
837 GAGTCTCATCACTCTCGCTCTACGCCACAGACAGAGTGTGGAGAC 886
234 euGlyLupPheHisAlaPheIleValIleSerAspLeuArgGluLeuGln 250
887 TGGGGAGTTCATGCTTCATTCTCTCGACCTCGGGAGCTGACAGAT 936
251 GlnAlaGlyGlnCysIleLeuLeuLeuArgIleGlnAsnProTrpGlyAr 267
937 CAGGGGGGCGAAGCATCTCTGCTGGAGTCCAGAACCCCTGGGGCGG 986
267 gArGcYsTrpGlnGlyLeuTrpArgGluGlyGlyGlyGlyTrpSerGln 284
987 GCGGTGCTGGCAGGGGCTCTGGAGAGGGGGGTGAAGGTGGAGCCAGG 1036
284 alaSPAlaAlaValAlaSerGluLeuLeuSerGlnLeuGlnGlyGln 300
1037 TAGATGCAAGCGGTACGTCGTAGCTCCCTGCCAGCTCCAGGAAGGGAG 1086
301 PheTrpValGluGluGluPheLeuArgGluPheAspGluLeuThrVa 317
1087 TTCTGGGTGGAGAGAGAGTTCCTCAGGAGATTGACAGATCACCAT 1136
317 lGlyTrpValThrGluAlaGlyHisLeuGlnSerLeuTrpThrGln 334
1137 TGGTACCAATCACAGAGGCGGCGACCTGCAGAGCTTACACAGGA 1186
334 rglLeuLeuCysHisThrArgAlaLeuProGlyAlaTrpValLysGlyGln 350
1187 AGCTGCTTGGCACACTCGGGCGCTGGGGCTGGGTCAAGGGGAG 1236
351 SerAlaGlyGlyCysArgAsnAsnSerGlyPheProSerAsnProLysPh 367
|||||
1237 TCAGCAGAGAGCTGCCGMAACAGCGGCTTCCACAGAACCCCAATT 1286
367 eTrpLeuArgValSerGluProSerGluValTyrIleAlaValLeuGln 384
1287 CTGGCTGGGGGTCTCAGAACCGAGTACATTCCCTCCCTGCAGAGA 1336
384 rGSerArGlnHisAlaIleAspTrpAlaGlyArgAlaArgAlaLeuVal 400
1337 GATCAAGCTGCGTCACTGAGTGGAGCGCGGGCCCGGGCACTGGTG 1386
401 GlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHisTy 417
1387 GGTGACAGTACATCTTCGTGGAGGCCACAGACATCCAGGCAACACTA 1436
417 rGlnAlaValGlyLeuHisLeuTrpLysValGluLysArgValAsnL 434
1437 CCAGGCTGGGCGCTGCACCTCTGGAGGTGGAGAACGCGGCGTCAATC 1486
434 euProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAla 450
1487 TCCCAGGGTCTTTCACAGCCCGCTGGCTGGACTGTGCCTGCACAGCA 1536
451 TyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTyrTyrLe 467
1537 TAGACCGGGAGGTCCACTGCTGGAGGCTTCACCGGGCTACTACT 1586
467 uAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeuLeu 484
1587 GCGTCTCCCGACACCTTCTGMAAGAGCGCCAGGGGAGTTCCTGCTCC 1636
484 rGValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAla 500
1637 GAGTCTTCTTACCGGGAGTCTCTTACGGCCATGAGGCAAGTGGCC 1686
501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyLupTrpGlyThrVa 517
1687 AAGAAGCGCACCCCGACAGAGCCCTGCTGGGGGAGTGGGGAGCCGT 1736
517 lGlnLeuArgGlySerTrpArgValGlyGlnThrAlaGlyLysSerArg 534
1737 GCAGCTGGGGGCTCTCTGAGAGCGCGACAGCGCGGGGGCAGACAGA 1786
534 snPheAlaSerTyrProThrAsnProCysPheProPheSerValProGln 550
1787 ACTTGGCTGTACCCCAACCAACCCCTGCTTCCCTTCTCGGTCGCCAG 1836
551 GlyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArgPr 567
1837 GGGCTGGCGCCCGGTGGCTGGCATGCTGTGATGACATGACATGCGAAC 1886
567 oSerAspThrLupPheHisProIleGlyPheHisIlePheGlnValProG 584
1887 CAGGACACCCAGTCCACCCCATCGGCTTCATATCTTCAGAGTCCAG 1936
584 lUGlyLysArgSerGlnAspAlaProProLeuLeuLeuGlnGluProLeu 600
1937 AGGCTGGAGAGCGCAGAGCGCGCCACTGTGCTGCAGAGACCGGTG 1986
601 LeuSerCysValProHisArgTyrAlaGlnGluValSerArgLeuCysLe 617
1987 CTGAGCTGCGCGCATCGGTATGCCCAGAGGTGAGCCAGCTGTGCT 2036
617 uLeuProAlaGlyThrTyrLysValValProSerThrTyrLeuProAsp 634
2037 CCTGCTTCCAGCACTACAGAGGTGTCCTCCACCTACCTGCAGACAA 2086
634 hrGluGlyAlaPheThrValThrIleAlaThrArgIleAspArgProSer 650
2087 CAGAGGGGGCTTCACTGTGACATCGCMAACAGATTTGACAGTTC... 2132
651 lIleHisSerGlnIleMetLeuGlyGlnPheLeuGlnGluValSerValMe 667
2133 .....TCCGTCAT 2140

```

667 talavalmetlythr 672
2141 GCCAGTGTAAAC 2156

seq_name: gb_pat:ARI53291

seq_documentation_block:

LOCUS ARI53291 2297 bp linear PAT 08-AUG-2001
DEFINITION Sequence 5 from patent US 6235481.
ACCESSION ARI53291
VERSION ARI53291.1 GI:15120823
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.

REFERENCE 1 (bases 1 to 2297)

AUTHORS Horikawa, Y., Oda, N., Hanis, C.L., Bell, G.I. and Cox, N.J.
TITLE Polynucleotides encoding calpain 10
JOURNAL Patent: US 6235481-A 5 22-MAY-2001;
FEATURES Location/Qualifiers
1..2297
/organism="unknown"

BASE COUNT 404 a 713 c 772 g 408 t
ORIGIN

alignment_scores:

Quality: 3290.00 Length: 672
Ratio: 5.350 Gaps: 1
Percent Similarity: 91.518 Percent Identity: 91.518

alignment_block:

US-09-768-877-2 x ARI53291

Align seg 1/1 to: ARI53291 from: 1 to: 2297

1 MetArgAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
|||||
42 ATGCGGG 91
17 aAlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrProL 34
|||||
92 CGGCTTCCCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 141
34 euAlaGlnPheArgGluAspIleThrTrpArgArgProGlnGluIleCys 50
|||||
142 TGGCCCCAGTTCCGGAGGACATCACGTGGAGCGGGCCCGCAGGAGTTTCT 191
51 AlaThrProArgLeuPheProAspAspProArgGluGluGlnVallysl 67
|||||
192 GCCACACCCCGGCTGTTCCAGATGACCACCGGAAGGCGAGGTGAAGCA 241
67 nGlyLeuLeuGlyAspCysTrpPheLeuCysAlaCysAlaAlaLeuGlnL 84
|||||
242 GGGGTGCTGGGGATTGCTGTTCTGTTCTGCTGCTGCTGCTGCTGCTG 291
84 ySerArgHisLeuLeuAspGlnValIleProProGlyGlnProSerTrp 100
|||||
292 AGAGCAGGCACCTCTCTGGACCAAGTATCTCCCTCCGGCAGCGAGCTGG 341
101 AlaAspGlnGluTyArgGlySerPheThrCysArgIleTrpGlnPheGl 117
|||||
342 GCGGACAGGAGTACCGGGGGCTCTTCCCTGCTGCTGCTGCTGCTGCTG 391
117 yArgTrpValGluValThrThrAspAspArgLeuProCysLeuAlaGlyA 134
|||||
392 ACGCTGGGTGGAGTGACACAGATGACCGCTGCTGCTGCTGCTGCTGCTG 441
134 rgLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
|||||
442 GACTCTGTTTCTCCCGCTGCCAGAGGAGGATGTTCTTGGCTCCCTTA 491

151 LeuGluLysValTyrAlaLysValHisGlySerTyrTyrGluHisLeuTrpAl 167
|||||
492 CTGGAAGAGGTCTACGCCAAGGTCCATGGGTCTACAGACACCTGTGGGG 541
167 aglyGlnValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGluA 184
|||||
542 CGGGCAGGTGGCGATGCGCTGTGTGACCTGACCGCGGCTGTGGCAGAA 591
184 rGTrpAsnLeuLysGlyValAlaGlySerGlyGlyGlnGlnAspArgPro 200
|||||
592 GATGGAACCTGAAGGCGGTAGCAGGAAGCGGAGCCAGGCGGACAGGCCA 641
201 GlyArgTrpGluHisArgThrCysArgGlnLeuLeuHisLeuLysAspGl 217
|||||
642 GGGCGCTGGGAGCACAGGACTTGTGGCAGCTGCTCCACCTGAAGGACCA 691
217 nCysLeuIleSerCysValLeuSerProArgAlaGlyAlaArgGluL 234
|||||
692 GTGTCTGTATCAGCTGCTGCTGCTCAGCCCGCAGAGCAGGTGCGCGGAG 741
234 euGlyGluPheHisAlaPheIleValSerAspLeuArgGluLeuGlnGly 250
|||||
742 TGGGGAGTTCCATGCTTCTGCGACCTGCGGAGCTGCGGAGCTCCAGGCT 791
251 GlnAlaGlyGlnCysIleLeuLeuLeuArgIleGlnAsnProTrpGlyAr 267
|||||
792 CAGGGCGGCAGTGCATCTGCTGCTGCGGATCCAGAACCTCTGGGCGCG 841
267 gArgCysTrpGlnGlyLeuTrpArgGluGlyGlyGlyTrpSerGlnV 284
|||||
842 GCGGTGCTGGCAGGGGCTCTGGAGAGAGGGGGTGAAGGCTGGAGCCAGG 891
284 aAlaPheAlaValAlaSerGluLeuLeuSerGlnLeuGlnGluGlyGlu 300
|||||
892 TAGATGACCGGTAGCATCTGAGCTCTGCTCCAGCTCCAGGAAGGGGAG 941
301 PheTrpValGluGluGluPheLeuArgGluPheAspGluLeuThrVa 317
|||||
942 TTCTGGGTGGAGGAGGAGGTCTCTCAGGAGTTTGGAGGACCTCACCT 991
317 lGlyTrpProValThrGluAlaGlyHisLeuGlnSerLeuThrGluA 334
|||||
992 TGGGTACCGGGTACGGAGGGCGGGCCACCTGCAGAGCCTCTACAGAGA 1041
334 rGLeuLeuCysHisThrArgAlaLeuProGlyValTrpValLysGlyGln 350
|||||
1042 GCCTGCTGCTGCTATACGGGGGCTGCTGCTGGGCTGGGTCAAGGGCCAG 1091
351 SerAlaGlyGlyCysArgAsnAsnSerGlyPheProSerAsnProLysPh 367
|||||
1092 TCAGCAGGAGGCTGCCGGAACAACAGCGCTTTCCACAGCAACCCCAATT 1141
367 eTrpLeuArgValSerGluProSerGluValTyrIleAlaValLeuGlnA 384
|||||
1142 CTGGCTGGGGTCTCAGAACCGAGTGAGTGTATATTGCGCTCTCTGCAGA 1191
384 rGSerArgLeuHisAlaAlaAspTrpAlaGlyArgAlaArgAlaLeuVal 400
|||||
1192 GATCCAGGCTGCACCGCGGAGCTGGGAGGGCGGGCGGGCGGCGCTGGTG 1241
401 GlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHisTy 417
|||||
1242 GGTGACATCATACTCTGTTGGAGCCAGCAGCATCCCGGGGCAAGCACTA 1291
417 rGlnAlaValGlyLeuHisLeuTrpLysValGluLysArgValAsnL 434
|||||
1292 CCAGGCTGTGGGTCTGCACCTCTGGAGGTAGAGAAGCGCGGGTCAATC 1341
434 euProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAla 450
|||||
1342 TGCTTAGGGTCTCTCCATGCCCCCGTGGCTGGCAGCCCGCTGCCATGCA 1391
451 TyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTyrTyrLe 467

```

|||||
1392 TACGACCGGAGGTCACCTGCTGTGAGCTCAGCCGCGCTACTACT 1441
467 uAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeu 484
|||||
1442 GCGTCTCCCGACCGCTCTCTAGAGACCGCGGAGGAGCTCCGCTCC 1491
484 rGValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaVal 500
|||||
1492 GAGCTTCTCTACCGGCGAGCTCCCTAG..... 1522
501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyLutPrgLys 517
1522 ..... 1522
517 LGlutLeuArgLysSerTrpArgValGlyGlnThrAlaGlySerArg 534
1522 ..... 1522
534 snPheAlaSerTyrProThrAsnProCysPheProPheSerValPro 550
1522 ..... 1522
551 GlyProGlyProArgCysValArgLleThrLeuHisGlnHisCysArg 567
|||||
1523 GGCCTGGCCCCCGCTCGCCGATCCTGCTCATCAGCACTGCGCGGCC 1572
567 oSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValPro 584
|||||
1573 CAGTGACACCGAGTCCACCCCATCGCTTCCATATCTTCCAGGCTCCAG 1622
584 LuGlyGlyArgSerGlnAspAlaProProLeuLeuGlnGluProLeu 600
|||||
1623 AGGCTGAGAGAGCCAGGACGACCCCATCGCTGCTCAGAGAGCGCTG 1672
601 LeuSerCysValProHisArgTyrAlaGlnGluValSerArgLeuCys 617
|||||
1673 CTGAGCTGCGGCACATCGCTACGCCAGAGGAGCGCGCTGCTGCT 1722
617 uLeuProAlaGlyThrTyrLysValValProSerThrTyrLeuProAsp 634
1723 CCGTCCGTCAGGACACTACAGGTTGTGCTCCACTCACTGCGCGGCA 1772
634 hrGluGlyAlaPheThrValThrIleAlaThrArgIleAspArgPro 650
|||||
1773 CAGAGGGGCGCTTCACAGTGCATCGCACAGAGATTGACAGCGCATCC 1822
651 IleHisSerGlnGluMetLeuGlyGlnPheLeuGlnGluValSerVal 667
1823 ATTTCAGACCGCAGGAGATGCTGGCCAGTTCCTCCAGAGGCTCCGTCAT 1872
667 tAlaValMetLysThr 672
1873 GCGAGTATGATAAAC 1888
seq_name: gb_dr:AF089090
seq_documentation_block:
LOCUS AF089090 2297 bp mRNA linear PRI 23-JAN-2001
DEFINITION Homo sapiens calpain-like protease CAPN10b mRNA, complete cds.
ACCESSION AF089090
VERSION AF089090.1 GI:10503938
KEYWORDS
SOURCE
ORGANISM Homo sapiens
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 2297)
REFERENCE
AUTHORS Horikawa,Y., Oda,N., Cox,N.J., Li,X., Orho-Melander,M., Hara,M.,
Hirokawa,Y., Lindner,T.H., Mashima,H., Schwarz,P.E., del
Bosque-Plata,L., Horikawa,X., Oda,Y., Yoshuchi,I., Collila,S.,
Polonsky,K.S., Wei,S., Concanon,P., Iwasaki,N., Schuitze,J.,
Baler,L.J., Bogardus,C., Groop,L., Boerwinkle,E., Hanis,C.L. and

```

```

TITLE
Genetic variation in the gene encoding calpain-10 is associated
with type 2 diabetes mellitus
JOURNAL Nat. Genet. 26 (2), 163-175 (2000)
MEDLINE 20472315
REFERENCE 2 (bases 1 to 2297)
AUTHORS Horikawa,Y. and Bell,G.I.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-1998) Howard Hughes Medical Institute, The
University of Chicago, 5841 S. Maryland Avenue, MC1028, Chicago, IL
60637, USA
FEATURES
source
1. 2297
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2q37"
42..1676
/note="type B: alternatively spliced"
/codon_start=1
/product="calpain-like protease CAPN10b"
/protein_id="A017968.1"
/db_xref="GI:10503939"
/translation="MRAGKATFPARELFDAAPADSSIFCDLSTPLAOFREDITWR
ROEICATPRLPDDPRGOVKGILGIDMFLCACAALQKSHLLDOVIIPROPSWAD
OYRSGSFTRIMQGRWVETDNDRLPCLAGRCFSCRCREDVFTLPLEKTYARVHG
STEHLAGOVADALVDLTGGLAEKRWMLKCVAGSGGQDGRGHEHRTQRLHLKQOC
LISCCVLSPPAGARELGEFHAFTVSDRLQAGQCLILLRIONPWRRCWGLMRE
GGEGWQVDAVAASBELSQLQEBEWEVEEELREFDELTVGYPTVEACHLSLYTER
LCHTRPALPGAWKQGSAGCRNNSGFPENPKFMLEVSESEYTLAVLQSRSLHADM
AGRAALVDSHRSNPSASIPGKHQAVGLHLMKVKRRVNLPRVLSMPVAGTACHA
YAREHLRCELSPGYLAIVSTFLDAPDELAPDELAPDELAPDELAPDELAPDELAP
ASPVTPSSRPSASISRSQREBARTHPCCRSRC"
BASE COUNT 404 a 713 c 772 g 408 t
ORIGIN
polya_signal
2289..2294
alignment_scores:
Quality: 3290.00 Length: 672
Ratio: 5.350 Gaps: 1
Percent Similarity: 91.518 Percent Identity: 91.518
alignment_block:
US-09-768-877-2 x AF089090 ..
Align seg 1/1 to: AF089090 from: 1 to: 2297
1 MetArgAlaGlyArgGlyAlaThrProAlaArgLysLeuPheArgAspAl 17
|||||
42 ATGCGGGCGGGCGGGCGGCGGCGGCGGAGGAGCTGTCCGGGAGCG 91
17 aAlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrPro 34
|||||
92 CGCCTCCCGCGCGCGGAGCTCTGCTCTTCTGAGACTGTGTCAGCGCGC 141
34 euAlaGlnPheArgGlyAspPheThrTrpArgArgProGlnGluIleCys 50
|||||
142 TGGCCGAGTTCGCGGAGACATCACGAGGAGCGCGCCAGACATTTGT 191
51 AlaThrProArgLysLeuPheProAspAspProArgGlyGlnValLysG 67
192 GCCACACCGCGCGCTGTTCAGATGACCCACGGAAGGCGAGGTGAAGCA 241
67 ngLysLeuGlyAspCysTrpPheLeuCysAlaCysAlaLeuGlnL 84
|||||
242 GGGCTCTCGGGGAGATTGCTGTTCCGTGTGCTGCGCGCGCTGCAGCA 291
84 ySerArgHisLeuLeuAspGlnValLleProProGlyGlnProSerTrp 100
|||||
292 AGAGCAGGACATCTCTGAGACAGGTCACTTCCGCGGACAGCGAGCTGG 341
101 AlaAspGlnGlyArgGlySerPheThrCysArgLleThrPrgLysPheG 117

```

|||||
342 GCCACGAGGAGTACCGGGGCTCTTCCACCTGCGCATTTGGCAGTTGG 391
|||||
117 YATrPValGluValThrThrAspArgLeuProCysLeuAlaGlyA 134
|||||
392 ACGCTGGTGGAGGTGACACAGATGACCGCTGCCGTGCTTGCAGGGA 441
|||||
134 rgLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
442 GACTCTGTTCTCCCGTGCAGAGGAGGATGTTCTGCTGCCCTCCCTTA 491
|||||
151 LeuGluLysValTyrAlaLysValHisGlySerTyrGluHisLeuTrpAl 167
492 CTGGAAGGTCACGCAAGGTCATGGGTCTTACGAGCACCTGTGGGC 541
|||||
167 aglyGlnValAlaAspAlaLeuValAspLeuThrGlyLeuAlaGluA 184
542 CGGCAGGTGGCGATGCCCTGGTGGACCTGACCGCGCGCTGGCAGAA 591
|||||
184 rgTrpAsnLeuLysGlyValAlaGlySerGlyGlnGlnAspArgPro 200
592 GATGGAACCTGAAGGGCTAGCAGGAAGCGAGCGCCAGCAGCAGGCCA 641
|||||
201 GlyArgTrpGluHisArgThrCysArgGlnLeuLeuHisLeuLysAspG 217
642 GGCCTGCTGGGAGCACAGGACTTGTGCGCAGCTGCTCCACCTGAAGGACCA 691
|||||
217 nCysLeuIleSerCysCysValLeuSerProArgAlaGlyAlaArgGluL 234
692 GTGCTGATCAGCTGTGCGGTGTGAGCCCGCAGCAGAGGTGCGCGGAGC 741
|||||
234 euGlyGluPheHisAlaPheIleValSerAspLeuArgGluLeuGlnGly 250
742 TGGGGAGTTCCATTCCTTCTGCGACCTGCGGGAGCTCCAGGGT 791
|||||
251 GlnAlaGlyGlnCysIleLeuLeuLeuArgIleGlnAsnProTrpGlyAr 267
792 CAGCGGGCCAGTGCATCTCTGCTGCTGCGGATCCAGAACCCCTGGGGCCG 841
|||||
267 gArCysTrpGlnGlyLeuTrpArgGluGlyGlyGlyTrpSerGlnV 284
842 CGGCTGCTGGCAGGGCTCTGGAGAGAGGGGGGTGAAGGTGGAGCCAGG 891
|||||
284 alAspAlaValAlaSerGluLeuLeuSerGlnLeuGlnGluGlyGlu 300
892 TAGATGACGGGTAGCATCTGAGCTCTGCTCCAGCTCCAGGAAGGGGAG 941
|||||
301 PheTrpValGluGluGluPheLeuArgGluPheAspCylLeuThrVa 317
942 TTCCTGGGTGGAGGAGGAGGAGTTCCCTCAGGGAGTTTGACGAGCTCACCGT 991
|||||
317 lGlyTyrProValThrGluAlaGlyHisLeuGlnSerLeuTyrThrGluA 334
992 TGCTACCCGGTCAAGGAGCGCGCCACCTGCGAGAGCCTCTACACAGAGA 1041
|||||
334 rgLeuLeuCysHisThrArgAlaLeuProGlyAlaTrpValLysGlyGln 350
1042 GGCTGCTCTGCCATACCGCGGGCTGCTGCTGGGCTGGGTCAAGGGCCAG 1091
|||||
351 SerAlaGlyCysArgAsnAsnSerGlyPheProSerAsnProLysPh 367
1092 TCAGCAGGAGGCTGCGCGGAACAACAGCGGCTTTCCAGCAACCCCAAT 1141
|||||
367 eTrpLeuArgValSerGluProSerGluValTyrIleAlaValLeuGlnA 384
1142 CTGGCTCGGGTCTCAGACCGAGTGGGTGTACATTGGCTCTCCAGCA 1191
|||||
384 rgSerArgLeuHisAlaAlaAspTrpAlaGlyArgAlaArgAlaLeuVal 400
1192 GATCCAGGCTGACGCGGGGAGTGGGCGCGCGCGCGCGCGCGCGCGCG 1241
|||||
401 GlyAspSerHisThrSerTrpProAlaSerIleProGlyLysHisTy 417
|||||

1242 GGTGACAGTCATCTTCTGGAGCCCGAGCATCCCGGCAAGCACTA 1291
417 rGlnAlaValGlyLeuHisLeuTrpLysValGluLysArgArgValAsnL 434
1292 CAGGCTGTGGGTCTGCACCTCTTGAAGGTAGAGAAGCGCGGTCAATC 1341
434 euProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAla 450
1342 TGCTAGGTGCTCTGTCATGTCCTCCCGTGGTGGACCGGCTCCATGCA 1391
451 TyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTyrTyrLe 467
1392 TAGCACCGGAGGTCCACCTGCTGTGAGCTCTCACCGGGCTACTACCT 1441
467 uAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeuLeuA 484
1442 GGTGTCTCTACCGGGCAGCTCTCCCTTAG..... 1491
484 rgValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAla 500
1492 GAGTCTCTACCGGGCAGCTCTCCCTTAG..... 1522
501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyGluTrpGlyThrVa 517
1522 1522
517 lGlnLeuArgGlySerTrpArgValGlyGlnThrAlaGlyGlySerArgA 534
1522 1522
534 snPheAlaSerTyrProThrAsnProCysPheProPheSerValProGlu 550
1522 1522
551 GlyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArgPr 567
1523 GGCCTTGGCCCGCTGCTGCTGCGCATCTCTGCATCAGCACTGCGCGCC 1572
567 oSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValProG 584
1573 CAGTGACACCGAGTTCCACCCCATCGGCTTCCATATCTTCCAGGTCCTAG 1622
584 luGlyGlyArgSerGlnAspAlaProProLeuLeuLeuGlnProLeu 600
1623 AGGTTGAAGAGGAGCAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1672
601 LeuSerCysValProHisArgTyrAlaGlnGluValSerArgLeuCysLe 617
1673 CTGAGCTGCTGCCATCTGCTACGCCAGGAGGTGAGCGGCTCTGCT 1722
617 uLeuProAlaGlyThrTyrLysValValProSerThrTyrLeuProAsp 634
1723 CTGCTGCGAGGACCTACAGGTTGTGCTTCCACCTACCTGCGCGGACA 1772
634 hrGluGlyAlaPheThrValThrIleAlaThrArgIleAspArgProSer 650
1773 CAGAGGGGCTTCCACAGTACCATCGCAACCCAGGATTCAGAGGCAATCC 1822
651 lIleHisSerGlnGluMetLeuGlyGlnPheLeuGlnGluValSerValMe 667
1823 ATTACAGCCAGGAGATGCTGGCGCAGTTCTTCCAAAGAGGTCTCGCTCAT 1872
667 tAlaValMetLysThr 672
1873 GGCAGTGATGAAANACC 1888

seq_name: gb_pat:AR153294

seq_documentation_block:

LOCUS AR153294

DEFINITION Sequence 11 from patent US 6235481.

ACCESSION AR153294

VERSION AR153294.1 GI:15120826

2516 bp

DNA

linear

PAT 08-AUG-2001


```

KEYWORDS      Unknown.
SOURCE        Unknown.
ORGANISM      Unclassified.
REFERENCE     1 (bases 1 to 2516)
AUTHORS       Horikawa, Y., Oda, N., Hants, C.L., Bell, G.I. and Cox, N.J.
TITLE        Polynucleotides encoding calpain 10
JOURNAL       Patent: US 6235481-A 11 22-MAY-2001;
FEATURES
  source      1..2516
              Location/Qualifiers
BASE COUNT   445 a 777 c 837 g 457 t
ORIGIN
alignment_scores:
  Quality: 3239.50      Length: 745
  Ratio: 5.267          Gaps: 2
  Percent Similarity: 82.350  Percent Identity: 82.416
alignment_block:
  US-09-768-877-2 x AR153294
Align seg 1/1 to: AR153294 from: 1 to: 2516

1 MetArgAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
  |||||||
42 ATGGGGCGGGCGCGGGCGGACCGCGGAGAGGTTCCTCCGGGAGCC 91
  |||||||
17 aaIaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrProL 34
  |||||||
92 CGGCTTCCTCCGCGCGGACCTCTGCTCTGCGACTTCTCTACGCCGC 141
  |||||||
34 euAlaGlnPheArgGluAspIleThrTrpArgArgProGlnGluIleCys 50
  |||||||
142 TGGCCGAGTTCGGGAGGACATCATCGTGAAGGCCGCCAGAGATTGTT 191
  |||||||
51 AlaThrProArgLeuPheProAspAspProArgGlnGluValIleYsgI 67
  |||||||
192 GCCACACCGCGGCTGTTCCAGATGACCCAGGAGGAGGAGGTGAAGCA 241
  |||||||
67 ngIleuLeuGlyAspCysTrpPheLeuCysAlaCysAlaAlaLeuGlnL 84
  |||||||
242 GGGGCTCTGGGGATTGCTGTGCTCTGTGCTGTGCGCGCGCTGCAGA 291
  |||||||
84 ySserArgHisLeuLeuAspGlnValIleProProGlnGlnProSerTrp 100
  |||||||
292 AGAGCAGAGACCTCTCTGACAGGTCTCTCTCCGGGACAGCCAGCTGG 341
  |||||||
101 AlaAspGlnGluTyrArgGlySerPheThrCysArgIleTrpGlnPheG 117
  |||||||
342 GCCGACACAGAGATACCGGGGCTCTCTCACTGTGTCATTTGGCAGTTGG 391
  |||||||
117 yArgTrpValGluValThrThrAspAspArgLeuProCysLeuAlaGly 134
  |||||||
392 ACCGTGGGTGAGGTGACACAGATGACCGCTGCTGCTTGCACAGGA 441
  |||||||
134 rglLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
  |||||||
442 GACTCTGTTTCTCCCGCTGCGCAGAGGAGATGTTCTGCTGCCCTTGA 491
  |||||||
151 LeuGlnIleValTyrAlaTyrValHisGlySerTyrGlnHisLeuTrpAl 167
  |||||||
492 CTGGAAGAAAGTCTACGCCAAGTCCAGAGGTCTCTACAGACACTGTGGC 541
  |||||||
167 agIyGlnValAlaAspAlaLeuValAspLeuThrGlyIleuAlaGluA 184
  |||||||
542 CGGGCAGGTGGCGGATGCCCTGCTGGACCTGACCGGGCGCTGGCAGAAA 591
  |||||||
184 rgrTrpAsnLeuIleYsgIValAlaGlySerClyGlyGlnGlnAspArgPro 200
  |||||||
592 GATGAAACCTGAAGGCGTAGCAGAGAGCGAGGCCAGCAGCAGAGCCCA 641

```

```

201 GLVArTrpGlnHisArgThrCysArgGlnLeuLeuHisLeuYsAspG 217
  |||||||
642 GGCCGCTGGAGCAGACAGACTTGTGGCAGCTGTCCACCTGAAAGACCA 691
  |||||||
217 nCysLeuIleSerCysCysValLeuSerProArgAlaGlyAlaArgGluL 234
  |||||||
692 GTGCTGTATCAGCTGCTGCGTGTCTACACCCAGAGCAGGTGCCGGAGC 741
  |||||||
234 euGlyGluPheHisAlaPheIleValSerAspLeuArgGluLeuGlnGly 250
  |||||||
742 TGGGGGAGTTCATCCTTCTGCTGTGACCTGCGGAGCTCCAGAGGT 791
  |||||||
251 GlnAlaGlnCysIleLeuLeuLeuArgIleGlnAsnProTrpGlyArg 267
  |||||||
792 CAGCGGGGCCAGTCCATCTGCTGTCTGCGATCCAGAACCCCTGGGCGC 841
  |||||||
267 gArgCysTrpGlnGlyLeuTrpArgGlnGlyGlyGlyTyrSerGlnV 284
  |||||||
842 GCGGTGCTGGCAGGGGCTGTGAGAGAGAGGGGGTGAAGGTGGAGCCAG 891
  |||||||
284 aIaAspAlaIaValAlaSerGluLeuLeuSerGlnLeuGlnGlyGlu 300
  |||||||
892 TAGATGACGGGTAGCATCTGAGCTCTGTCCTCCAGCTCCAGAGGGAG 941
  |||||||
301 PheTrpValGlnGlnGlnGlnGlnPheLeuArgGlnPheAspGluLeuThrVa 317
  |||||||
942 TTCGGGTGAGAGAGAGAGAGAGTTCCTCAGGAGATTATACAGAGCTCACC 991
  |||||||
317 IgIYTrpProValThrGlnAlaGlnHisLeuGlnSerLeuTyrThrGluA 334
  |||||||
992 TGCGTACCGCGTACAGGAGCGCGGCACCTGACAGAGCTCTACACAGAGA 1041
  |||||||
334 rglLeuLeuCysHisThrArgAlaLeuProGlyAlaTrpValIleYsgI 350
  |||||||
1042 GGGCTCTCTCCATACCGGGGCGCTGCTGGGCTGGGTCAAGGGCGCAG 1091
  |||||||
351 SerAlaGlyIleCysArgAsnAsnSerGlyPheProSerAsnProIlySph 367
  |||||||
1092 TCAGCAGAGAGCTGCCGGAACACAGGGGCTTTCACGACACCCAAATT 1141
  |||||||
367 eTrpLeuArgValSerGluProSerGluValTyrIleAlaValLeuGlnA 384
  |||||||
1142 CTGGCTGCGGGTCTCAAAACCGAGTGAAGTGTACATGCGCTCTGCAGA 1191
  |||||||
384 rgsErArgLeuHisAlaAlaAspTrpAlaGlyArgAlaAlaLeuVal 400
  |||||||
1192 GATCCAGGCTGCACGCGCGGACTGGGACGCGCGCGCGGCACTGGTG 1241
  |||||||
401 GlyAspSerHisThrSerTrpSerProAlaSerIleProGlyHisIleS 417
  |||||||
1242 GGTGACAGTATATCTCTGTGGAGCCCAAGCGATCCCGGGCAAGCACTA 1291
  |||||||
417 rGlnAlaValGlyLeuHisLeuTrpIys 426
  |||||||
1292 CCAAGCTGTGGTCTGACACTTGGAAGGTGTGAACCTTGGACACACAC 1341
  |||||||
426 ..... 426
1342 TGTTCCTGCTCCCTTATGATGATGTGGCCACATGATGTTCTTCTCT 1391
  |||||||
426 ..... 426
1392 GCAGAAAGAGTCTGTGAAGGCCCACTGTCCAGCAGCCCCAGAGTTGCT 1441
  |||||||
426 ..... 426
1442 GGGCCAGGTCCTTTGTGGGCCCACTACAGAGAGACTTGCAGGCTGC 1491
  |||||||
427 .....Val 427
1492 TGTCTGGACAGATACTGGCGCAGGAGCCAAAGTGAAGCCGGGATTTGTA 1541
  |||||||
428 GlnIysArgArgValAsnLeuProArgValLeuSerMetProProValAl 444

```

```

|||||
1542 GAGAGCGCGGGTCAATCTGCTAGGCTCTGTCCATGCCCGCGTGGC 1591
444 aGlyThrAlaCysHisAlaTyrAspArgGluValHisLeuArgCysGluL 461
|||||
1592 TGGCACCGCGTCCCATGATAGACCGGGAGGTCCACCTGCTGTGTGAGC 1641
461 euSerProGlyTyrTyrLeuAlaValProSerThrPheLeuLysAspAla 477
1642 TCTCACCAGGCTACTACCTGGCTGTCCACAGACCTTCTCTGAGGACGCG 1691
478 ProGlyGluPheLeuLeuArgValPheSerThrGlyArgValSerLeuSe 494
1692 CCAGGGAGTTCTCTCGAGTCTTCTACCGGGCGAGTCTCCCTTAG 1740
494 tAlaIleArgAlaValAlaLysAsnThrThrProGlyAlaAlaLeuProA 511
1741 ..... 1741
511 laGlyGluTyrPheValGlnLeuArgGlySerTrpArgValGlyGln 527
1741 ..... 1741
528 ThrAlaGlyGlySerArgAspPheAlaSerTyrProThrAsnProCysPh 544
1741 ..... 1741
544 eProPheSerValProGluGlyProGlyProArgCysValArgIleThrL 561
1742 ..... GGCCTGGCGCCCGCTGCTCGCTCGCATCACTC 1772
561 euHisGlnHisCysArgProSerAspThrGluPheHisProIleGlyPhe 577
1773 TGCATCAGCACTCGCGGCCAGTGACACCGAGTTCCACCCCATCGGCTTC 1822
578 HisIlePheGlnValProGluGlyClyArgSerGlnAspAlaProLeu 594
1823 CATATCTTCCAGGTCCCGAGGGTGGAGGAGCCAGGACGCCACCCCACT 1872
594 uLeuLeuGlnGluProLeuLeuSerCysValProHisArgTyrAlaGlnG 611
1873 GCTGTGCAGGAGCGCTGCTGAGCTGGTGGCCACATGCTACGCCCAGG 1922
611 luValSerArgLeuCysLeuLeuProAlaGlyThrTyrLysValValPro 627
1923 AGGTGAGCGCGTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1972
628 SerThrTyrLeuProAspThrGluClyAlaPheThrValThrIleAlaTh 644
1973 TCCACTACCTGCTGGACACAGAGGGGCTTCACAGTGACCATCGCAAC 2022
644 rArgIleAspArgProSerIleHisSerGlnGluMetLeuGlyGlnPheL 661
2023 CAGGATTCACAGGCCATCATTCACAGCCAGGAGATGCTGGGCCAGTTCC 2072
661 euGlnGluValSerValMetAlaValMetLysThr 672
2073 TCCAGAGGCTCTCGCTCATGTCAGTGATGATAAACC 2107
seq_name: gb_pr:AF089093

```

```

seq_documentation_block:
LOCUS AF089093 2516 bp mRNA linear PRI 23-JAN-2001
DEFINITION Homo sapiens calpain-like protease CAPN10e mRNA, complete cds.
ACCESSION AF089093
VERSION AF089093.1 GI:10503945
KEYWORDS
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 2516)
AUTHORS Horikawa,Y., Oda,N., Cox,N.J., Li,X., Orho-Melander,M., Hara,M.,

```

Hinokio,Y., Lindner,T.H., Mashima,H., Schwarz,P.E., del Bosque-Plata,L., Horikawa,Y., Oda,Y., Yoshiuchi,I., Colilla,S., Polonsky,K.S., Wei,S., Concannon,P., Iwasaki,N., Schuizer,J., Baier,L.J., Bogardus,C., Groop,L., Boerwinkle,E., Hanis,C.L. and Bell,G.I.
Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus
Nat. Genet. 26 (2), 163-175 (2000)
20472315
2 (bases 1 to 2516)
Horikawa,Y. and Bell,G.I.
Direct Submission
TITLE
Submitted (02-SEP-1998) Howard Hughes Medical Institute, The University of Chicago, 5841 S. Maryland Avenue, MC1028, Chicago, IL 60637, USA

FEATURES
source
1. .2516

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2q37"
42. .1376

CDS
/note="type E; alternatively spliced"
/codon_start=1
/product="calpain-like protease CAPN10e"
/protein_id="AAG17971.1"
/db_xref="GI:10503946"

/translation="MRAGRGATPARELFRDAAPPAADSLFCDLSTPLAQFREDITWR
RFQEICATPRLEPDDPRGQVKGGLGDCWFLCACAALQKSRHLLDQVLPKPPQSWAD
QYRGSFTCRIMQFGRVETDDRLPCLAGRLCFSRCOREDFVLPLEKYAVKHG
SYEHLWAGQVADALVDLTGLAERNLKGAGSGQDQGRWEHRTCOLLHLKDOC
LISCCVLPFRAGARELGEPAFIVSDRLQAGQOCILLRIQNPWGRRCQGLWRE
GCGMSQVDAAVASSELLSQLEGEFWEERLEPDELTVGYPTGAGHLASLTYER
LLCHTRALFPAWVKVQSGGCRNNSGFPFRFLVRKSPSEYITAVLQSRSLHAADM
AGRARALVGDSTHSWSPAPIKGVQAVGLHLMKGVTLGTLTFPPVPMWPT"

BASE COUNT 445 a 777 c 837 g 457 t
ORIGIN

alignment_scores:

Quality: 3239.50 Length: 745
Ratio: 5.267 Gaps: 2
Percent Similarity: 82.550 Percent Identity: 82.416

alignment_block:

US-09-768-877-2 x AF089093

Align seg 1/1 to: AF089093 from: 1 to: 2516

```

1 MetArgAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
|||||
42 ATGCGGCGGCGCGGCGCGGCGCGCGCGGCGGAGGCTGTTCGGGACGC 91
|||||
17 aAlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrProL 34
|||||
92 CGCCTTCCCGCGCGGAGCTCTGCTCTTCTGCGACTTGTCTACGCCGC 141
|||||
34 euAlaGlnPheArgGluAspIleThrTrpArgArgProGlnGluIleCys 50
|||||
142 TGGCCCACTTCCCGGAGCATCATCAGTGAGCGGCCGCCAGGAGATTGT 191
|||||
51 AlaThrProArgLeuPheProAspAspProArgGluGlyGlnValLysGl 67
|||||
192 GCCACACCCCGGCTGTTCAGATGACCCAGGAGGCGGAGGTTGAAGCA 241
|||||
67 nGlyLeuLeuGlyAspCysTrpPheLeuCysAlaCysAlaAlaLeuGlnL 84
|||||
242 GGGGCTCTCGGGGATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 291
|||||
84 ysSerArgHisLeuLeuAspGlnValIleProGlyGlnProSerTrp 100
|||||
292 AGAGCAGGACCTCTCGGACCATCAGGTCATTCTCTCGGACACAGCCG 341

```

101 AlaAspGlnGluTrpArgGlySerPheThrCysArgIleTrpGlnPheGln 117
 |||||
 342 GCCGACCAAGAGTACCGGGGCTCTTCACTCTCTCCATTGGCACTTTGG 391
 117 ValGTrpValGluValThrThrAspAspArgLeuProCysIleuAlaGlyA 134
 |||||
 392 ACCTGGGGTGGAGTGACACAGATGACCGCTGGCTGGCTTGCAGAGA 441
 134 rGLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
 |||||
 442 GACTCTGTTTCTCCGCTCCAGAGAGAGATGTTCTTGCTGCCCTTA 491
 151 LeuGluValValTyrAlaValHisGlySerTyrGlnHisLeuTrpAl 167
 |||||
 492 CTGGAAAAGGCTTACGCCACAGTCCATGGTCTACGACACCTGGGCG 541
 167 aGlyValValAlaAspAlaLeuValAspLeuThrGlyLeuAlaGln 184
 |||||
 542 CGGGCAGGTGGCGGATGCCCTGGTGGACCTGACGGCGGCTGGCAGAAA 591
 184 rGTrpAsnLeuLysGlyValAlaGlySerGlyGlnGlnAspArgPro 200
 |||||
 592 GATGGAACTGAGAGGCGCTACGAGAAACGGACGACAGACAGGCGCA 641
 201 GlyArgTrpGlnHisArgThrCysArgGlnLeuLeuHisLeuLysAspGln 217
 642 GCGCGCTGGAGCACAGGACTTGTGGCAGCTGCTCCACTGAAGAGACA 691
 217 nCysLeuLeuSerCysCysValLeuSerProArgAlaGlyAlaArgGlu 234
 |||||
 692 GTGCTGTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 741
 234 euGlyGlnPheHisAlaPheIleValSerAspLeuArgGlnLeuGlnGly 250
 |||||
 742 TGGGGAGATTCATGCCCTTCTCTCTGCGACCTCGGGAGCTCAGGGT 791
 251 GlnAlaGlyGlnCysIleLeuLeuLeuArgIleGlnAsnProTrpGlyArg 267
 |||||
 792 CAGCGGGGCGCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 841
 267 gArgCysTrpGlnGlyLeuTrpArgGlnGlyGlyGlyGlyTrpSerGln 284
 |||||
 842 GCGGTGCTGGCAGGGGCTCTGGAGAGAGGGGGTGAAGGGTGAAGCAGG 891
 284 AlaAspAlaAlaValAlaSerGlnLeuLeuSerGlnLeuGlnGlnGln 300
 |||||
 892 TAGATGACAGGGTGCATCTGAGCTCTCTCTCTCTCTCTCTCTCTCTCT 941
 301 PheTrpValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 317
 |||||
 942 TTTCTGGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 991
 317 IGLTyrProValThrGlnAlaGlyHisIleGlnSerLeuTyrThrGln 334
 |||||
 992 TGGCTACCGGTCACGAGAGCGGGCCACCTGACAGAGCTTACACAGAGA 1041
 334 rGLeuLeuCysHisThrArgAlaLeuProGlyAlaTrpValLysGln 350
 |||||
 1042 GCGTGGCTCTCCATACCGCGGGCTGCTGGGCTGGGCTGGGCTGGGCTG 1091
 351 SerAlaGlyLysCysArgAsnAsnSerGlyPheProSerAsnProLysPhe 367
 |||||
 1092 TCAGCAGAGAGGTCGCGGAACAACAGCGGCTTCCAGCAACCCCAATTT 1141
 367 eTrpLeuArgValSerLysProSerGlnValTyrIleAlaValLeuGln 384
 |||||
 1142 CTGGCTGGCGGCTCAAAACCGAGTGCAGTACATTGCCGCTCTGAGAGA 1191
 384 rGSerArgLeuHisAlaAlaAspTrpAlaGlyArgAlaArgAlaLeuVal 400
 |||||
 1192 GATCCAGCGCTGCACGCGCGGACTGGGACGCGGGCGCGGCGCTGGTG 1241
 401 GlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHisTyr 417

1242 GGTGACATCATCTACTTCTGTGAGGCCAGGAGCATCCGGGCAAGCACTA 1291
 417 rGlnAlaValAlaGlyLeuHisLeuTrpLys 426
 1292 CCAGGCTGTGGTCTGCACCTCTGTGAAGGGTGTGACACTTGGACACAC 1341
 426 426
 1342 TGTTCCTGTCCCTTCATGATGTGGCCACATGATGTTCCTTCCTT 1391
 426 426
 1392 GCMAAAGAGTTGGTGGAGGCCACACTGTCCAGACGCCCCAGGTGGCT 1441
 426 426
 1442 GGGCCACGTCCTTTTGGGCCAGCTACAAAGAGAGACTTGCAGGGCTG 1491
 427 427
 1492 TGTCTGGACAGATACCTGGCCGAGGCCAAGTAAAGCCCGGAGATTGGTA 1541
 428 GlnLysArgArgValAsnLeuProArgValLeuSerMetProProValAl 444
 1542 GAGAAGCGCGGGTCAATCTCTCTAGGGTCTGTCTCATGCCCCCGTGGC 1591
 444 aGlyThrAlaCysHisAlaTyrAspArgGluValHisLeuArgCysGln 461
 1592 TGGCAGCGCGTGCATGATGACCGGAGAGTGCACCTGTGGTGTGAGC 1641
 461 euserProGlyTyrTyrLeuAlaValProSerThrPheLeuLysAspAla 477
 1642 TCTACCGGGCTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1691
 478 ProGlyGlnPheLeuLeuArgValPheSerThrGlyArgValSerLeu 494
 1692 CCAGGGGAGTTCCTGCTCCGAGCTTCTCTACCGGGGAGTCTCTCTAG 1740
 494 rAlaIleArgAlaValAlaLysAsnThrThrProGlyAlaAlaLeuProA 511
 1741 1741
 511 IagGlyGlnTrpGlyThrValGlnLeuArgGlySerTrpArgValGlyGln 527
 1741 1741
 528 ThrAlaGlySerArgAsnPheAlaSerTyrProThrAsnProCysPhe 544
 1741 1741
 544 eProPheSerValProGlyGlyProGlyProArgCysValArgIleThrL 561
 1742 1742
 561 euHisGlnHisCysArgProSerAspThrGlnPheHisProIleGlyPhe 577
 1773 TGCATTCAGACTGCGGCCAGTGCACACGAGTTCACCCCATCGGCTTC 1822
 578 HisIlePheGlnValProGlyGlyArgSerGlnAspAlaProPhe 594
 1823 CATATCTTCCAGGTCCTCAGAGGGTGGAGAGGAGGAGGAGGAGGAGGAG 1872
 594 uLeuLeuGlnGlnProLeuLeuSerCysValProHisArgTyrAlaGln 611
 1873 GCTGCTCAGAGAGCGGCTGCTGAGCTGCTGCTGCTGCTGCTGCTGCTG 1922
 611 lValSerArgLeuCysLeuLeuProAlaGlyThrTyrLysValValPro 627
 1923 AGGTGACCGGCTCTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1972
 628 SerThrTyrLeuProAspThrGlnGlyAlaPheThrValThrIleAsn 644

434 euproaryValleuSerMetProProValAlaGlyThrAlaCysHisAla 450
 1342 TCCTAGAGGCTCTCCATGATGCCCCGCTGCGGACCCGCTGATGCA 1391
 451 TTTASPARGLuValHisLeuArgCysGluLeuSerProGlyTyrTyrLeu 467
 1392 TACGACCCGGAGGTCACCTGCTGTGAGCTCTCACCGGGCTACTACT 1441
 467 uAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeu 484
 1442 GGCTGTCCACACCTCTCTCAAGAGACCCGACGGGGATCTCTGCTCC 1491
 484 rGValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAla 500
 1492 GAGTCTTCTCTACCGGGCGAGCTCTCCCTAG..... 1522
 501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyGluTyrGlyThrVa 517
 1522 1522
 517 IGlnLeuArgGlySerTyrArgValGlyGlnThrAlaGlyGlySerArg 534
 1522 1522
 534 snPheAlaSerTyrProThrAsnProCysPheProPheSerValProGlu 550
 1522 1522
 551 GlyProGlyProAlaGlyCysValArgIlePheLeuHisGlnHisCysArgPr 567
 1522 1522
 567 oSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValProG 584
 1523 GTCCAG 1529
 584 IuGlyGlyArgSerGlnAspAlaProProLeuLeuGlnGluProLeu 600
 1530 AGGTGTGAGAGAGCCAGGACCCACCCGCTGCTGCTGAGAGCCGCTG 1579
 601 LeuSerCysValProHisArgTyrAlaGlnGluValSerArgLeuCysLe 617
 1580 CTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1629
 617 uLeuProAlaGlyThrTyrLysValValProSerThrTyrLeuProAspT 634
 1530 CCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1679
 634 hrGluGlyAlaPheThrValThrIleAlaThrArgIleAspArgProSer 650
 1680 CAGAGGGGGCTTCAAGTACATCGCAACAGGATGACAGGCAATCC 1729
 651 IleHisSerGlnLumMetLeuGlyGlnPheLeuGlnGluValSerValMe 667
 1730 ATTACAGCCAGAGATGCTGGGCGAGTCTCTCAAGAGGCTCTCCGTCAT 1779
 667 tAlaValMetLysThr 672
 1780 GGCAGTATGATAAAC 1795

seq_name: gb-pr:AF089092

seq_documentation_block:

LOCUS AF089092 2204 bp mRNA linear PRI 23-JAN-2001
 DEFINITION Homo sapiens calpain-like protease CAPN10d mRNA, complete cds.
 ACCESSION AF089092
 VERSION AF089092.1 GI:10503943

KEYWORDS

SOURCE

ORGANISM human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 2204)

AUTHORS Horikawa, Y., Oda, N., Cox, N.J., Li, X., Orho-Melander, M., Hara, M., Hinokio, Y., Lindner, T.H., Mashima, H., Schwarz, P.E., del Bosque-Plata, L., Horikawa, Y., Oda, Y., Yoshinuchi, I., Collilla, S., Polonsky, K.S., Wei, S., Concanon, P., Iwasaki, N., Schülze, J., Baler, L.J., Bogardus, C., Groop, L., Boerwinkle, E., Hanis, C.L. and Bell, G.I.
 TITLE Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus
 JOURNAL Nat. Genet. 26 (2), 163-175 (2000)
 MEDLINE 20472315
 REFERENCE 2 (bases 1 to 2204)
 AUTHORS Horikawa, Y. and Bell, G.I.
 TITLE Direct Submission
 JOURNAL Submitted (02-SEP-1998) Howard Hughes Medical Institute, The University of Chicago, 5841 S. Maryland Avenue, MC1028, Chicago, IL 60637, USA

FEATURES
 source 1..2204
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="2q37"
 42..1583
 /note="type D: alternatively spliced"
 /codon_start=1
 /product="calpain-like protease CAPN10d"
 /protein_id="A617970.1"
 /db_xref="GI:10503944"
 /translation="MRAGKATPARELEFRDAPRADSSLFCDLSPIAGFREDITMR
 RPOICATPRLEFPDDPRRGVYKGLIDCFPLACALAKSRHLLDQVTPRPPSRAD
 OERYGSETCINPGFQWVETTTDRPLCLAGRCFSRQREDYFMLEKVTAKVHG
 SYELWAGVADALVDLTGLARMLKGVAGSGGQODRGRNHRCTRQLILKQDC
 LISCVLSPVAGARELEFALFTVSDRELQAGQOCILLRIONPWRGRMGLMRE
 GGEWMSQVDAVAVSELSQLQDGEFVVEEERFDELTVGYPTAEAGHLSLTER
 LICHTRALPAMVKGOSAGGCRNNSGFPSPKFMRLSERSEVYIIVLQSRRLHADW
 AGRARLVGDSHTSWSPASIPCKHYQAVGILMLKVKRRVNIPLRVISMPVACTACHA
 YDREVLRLKELSGYLLAVPSFTLKPAPGFLRLRVSTGKVSILRSQREKARTHPCC
 CRRC"

BASE COUNT 390 a 672 c 752 g 390 t
 ORIGIN

alignment_scores:

Quality: 3090.50 Length: 672
 Ratio: 5.292 Gaps: 1
 Percent Similarity: 86.905 Percent Identity: 86.905

align_block:
 us-09-768-877-2 x AF089092 ..

Align seg 1/1 to: AF089092 from: 1 to: 2204

1 MetArgAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
 42 ATGG 91
 17 AlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrProL 34
 92 CGCCTTCCCGCCCGGACCTCTGCTCTTCTGCGACTTGTCTACGCCGC 141
 34 euAlaGlnPheArgGluAspIleThrThrPargArgProGlnGlnIleCys 50
 142 TGGCCAGATTCCGAGAGACATCAGTGGAGCGGCCGCCAGAGATTGTGT 191
 51 AlaThrProArgLeuPheProAspAspProArgGlnGlnValLysG1 67
 192 GCCACACCCCGGCTCTTCCAGATGACCCACGGAGAGGCGAGTGAAGA 241
 67 nGlyLeuGlnGlyAspCysTyrPheLeuCysAlaCysAlaAlaLeuGlnL 84
 242 GGGGCTGTGGGGGATGCTGCTCTCTGTCGTGGCGCCGCTGGAGA 291
 84 ySSerArgHisLeuLeuAspGlnValIleProProGlyGlnProSerTyr 100

|||||
292 AGACAGGACCTCTCGGACAGGTCAATCTCCGGACACCGGAGCTGG 341
|||||
101 AlaAspGlnGluTyrArgGlySerPheThrCysArgIleTrpGlnPheG1 117
|||||
342 GCCACACAGAGTACCGGGGCTCTTTCACCTGTGCGCATTTGGCAGTTGG 391
|||||
117 YArgTrpValGluValThrThrAspAspArgLeuProCysLeuAlaGlyA 134
|||||
392 ACGCTGGTGGAGGTGACACACAGATACCGCTGCCGTGCTTGCAGGGA 441
|||||
134 rGluCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
|||||
442 GACTCTGTTCTCCCGTCCAGAGGAGGATGTGTTCTGGCTCCCGCTTA 491
|||||
151 LeuGluLysValTyrAlaLysValHisGlySerTyrGluHisLeuTrpAl 167
|||||
492 CTGGAAGAGTCTACGCCAAGGTCCATGGGTCTTACGAGCACTGTGGGC 541
|||||
167 aGlyGlnValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGlu 184
|||||
542 CGGGCAGGTGGCGATGCTCGTGGACCTTGACCGCGCGCTGGCAGAAA 591
|||||
184 rGTrpAsnLeuLysGlyValAlaGlySerClyGlyGlnGlnAspArgPro 200
|||||
592 GATGGAACTTGAAGGGGTAGCAGGAAGCGGAGGCCAGCAGGACGCA 641
|||||
201 GlyArgTrpGluHisArgThrCysArgGlnLeuLeuHisLeuLysAspG1 217
|||||
642 GGCGCTGGGAGCACAGGACTTCTGGCAGCTGCTCCACCTGAAGGACCA 691
|||||
217 nCysLeuIleSerCysValLeuSerProArgAlaGlyAlaArgGluL 234
|||||
692 GTGCTGATCAGCTGTGCTGTCTAGCCAGCAGAGGTGCCCGGAGC 741
|||||
234 euGlyGluPheHisAlaPheIleValSerAspLeuArgGluLeuGlnGly 250
|||||
742 TGGGGAGTTCATGCTCTCATTTGTCTCGACCTGCGGGAGTCCAGGGT 791
|||||
251 GlnAlaGlyGlnCysIleLeuLeuLeuArgIleGlnAsnProTrpGlyAr 267
|||||
792 CAGCGCGGCAGTGCATCTGCTGCTGCGGATCCAGAACCCCTGGGGCGG 841
|||||
267 gArgCysTrpGlnGlyLeuTrpArgGluGlyGlyGlyGlyTrpSerGlnV 284
|||||
842 GCGGTCTGGCAGGGGCTCTGGAGAGGGGGGTGAAGGTGGAGCCAGG 891
|||||
284 aAlaAlaAlaValAlaSerGluLeuLeuSerGlnLeuGlnGluGlyGlu 300
|||||
892 TAGATCAGCGGTAGCATCTGAGCTCTCTGCCCAGCTCCAGGAAGGGGAG 941
|||||
301 PheTrpValGluGluGluGluPheLeuArgGluPheAspGluLeuThrVa 317
|||||
942 TTCTGGGTGGAGGAGGAGGAGTTCCTCAGGGAGTTTGACAGCTCACCGT 991
|||||
317 lGlyTyrProValThrGluAlaGlyHisLeuGlnSerLeuTyrThrGluA 334
|||||
992 TGCTTACCGGGTCACGGAGGGCGGCCACCTGCAGAGCCCTTACACAGAGA 1041
|||||
334 rGluLeuCysHisThrArgAlaLeuProGlyAlaTrpValLysGlyGln 350
|||||
1042 GGTGCTCTGCCATACGCGGGGCGCTGCTTGGGGGCTGGGTCAAGGGCCAG 1091
|||||
351 SerAlaGlyGlyCysArgAsnAsnSerGlyPheProSerAsnProLysph 367
|||||
1092 TCAGCAGGAGGCTGCCGGAACAACAGGCGCTTCCAGCAACCCCAATT 1141
|||||
367 eTrpLeuArgValSerGluProSerGluValTyrIleAlaValLeuGlnA 384
|||||
1142 CTGGCTGCGGGTCTCAGAACCGAGTGAGGTGTACATTGCCGTCTGCAGA 1191
|||||
384 rGSerArgLeuHisAlaAlaAspTrpAlaGlyArgAlaArgAlaLeuVal 400
|||||

1192 GATCCAGGCTGCACGGCGGAGCTGGGACGGCGGGCGGCGACCTGGTG 1241
|||||
401 GlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHisTy 417
|||||
1242 GGTGACAGTCATATCTTGTGGAGCCAGCAGCATCCCGGCAAGCACTA 1291
|||||
417 rGlnAlaValGlyLeuHisLeuTrpLysValGluLysArgArgValAsnL 434
|||||
1292 CCAGGCTGTGGGTCTGCACCTCTGGAAGGTAGAGACGGCGGGTCAATC 1341
|||||
434 euProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAla 450
|||||
1342 TGCTAGGGTCTGTCCATGCCCGCGTGGTGGCACCGCGTGCATGCA 1391
|||||
451 TyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTyrTyrLe 467
|||||
1392 TAGACCGGGAGGTCCACCTGTGCTGTGAGCTCTCACCGGGCTACTACCT 1441
|||||
467 uAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeuLeuA 484
|||||
1442 GGCTGTCCCAGCACCTTCTGAGGAGCGCCAGGGGAGTTCTGCTCC 1491
|||||
484 rGValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAla 500
|||||
1492 GAGTCTTCTTACCGGGGAGGTCTCCCTTAG 1522
|||||
501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyGluTrpGlyThrVa 517
|||||
1522 1522
|||||
517 lGlnLeuArgGlySerTrpArgValGlyGlnThrAlaGlyGlySerArgA 534
|||||
1522 1522
|||||
534 snPheAlaSerTyrProThrAsnProCysPheProPheSerValProGlu 550
|||||
1522 1522
|||||
551 GlyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArgPr 567
|||||
1522 1522
|||||
567 oSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValProG 584
|||||
1523 GTCCAG 1529
|||||
584 luGlyGlyArgSerGlnAspAlaProProLeuLeuLeuGlnGluProLeu 600
|||||
1530 AGGGTGAAGAGGAGCCAGGAGCGCACCCCTGCTGCTGCAGGAGCGCTG 1579
|||||
601 LeuSerCysValProHisArgTyrAlaGlnGluValSerArgLeuCyste 617
|||||
1580 CTGAGCTGCTGCCATCGCTACGCCAGGAGGTGAGCGGCTCTGCT 1629
|||||
617 uLeuProAlaGlyThrTyrLysValValProSerThrTyrLeuProAspT 634
|||||
1630 CCTGCTGCAGGACCTACAGGTGTGCTTCCCTTCCACCTACCTGCGGACA 1679
|||||
634 hrGluGlyAlaPheThrValThrIleAlaThrArgIleAspArgProSer 650
|||||
1680 CAGAGGGGCTTTCAGTGCACCTGCACCAACAGGATTGCACAGGCATCC 1729
|||||
651 ileHisSerGlnGluMetLeuGlyGlnPheLeuGlnGluValSerValMe 667
|||||
1730 ATTACAGCCAGGAGATGCTGGGCCAGTTCTTCCAAAGAGGTCTCGTCA 1779
|||||
667 tAlaValMetLysThr 672
|||||
1780 GCCAGTGATGAACACC 1795
|||||

seq_name: gb_010969
seq_documentation_block:

LOCUS BC010969 2528 bp mRNA linear ROD 25-JUL-2001
 DEFINITION Mus musculus, calpain 10, clone MGC:13731 IMAGE:4159612, mRNA,
 complete cds.
 ACCESSION BC010969
 VERSION BC010969.1 GI:15012121
 KEYWORDS MGC.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 Strausberg, R.
 Direct Submission
 Submitted (23-JUL-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 CONTACT MGC help desk
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
 contact: amadansystemsbiology.org
 Anup Madan, Rachel Dickhoff, Jessica Fahney, Stephanie Ford, Julia
 Greene, Mark Kettman and Anuradha Madan

REMARK
 COMMENT

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Series: IRK Plate: 18 Row: 1 Column: 9
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 6753259.

FEATURES

Source
 1. 2528
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="MGC:13731 IMAGE:4159612"
 /tissue="type=Liver, normal. 5 month old male mouse."
 /clone_1lb="NCI_CGAP_119"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 111..2111
 /codon_start=1
 /product="calpain 10"
 /protein_id="AAH10969.1"
 /db_xref="GI:15012122"
 /translation="MRAVRAETPARELPDAAFPASDSSILFNLSTPLAOPREDITMR
 RROEICATPOLFPDNDNMGVQGLQDCWFLCAQALQSOHLIDVQPPROGSGSD
 OKYGFETCRIMQFHWMEVTTIDRLPCLAGRLCSRCRQREDEYFRLPLEKATYAVHG
 STEHLMAGVADALVDLTGSLAEKRSLSKDYTAASGQDDPSGSEHRTQQLHLAKRC
 LISCVSLSPPAGARELGEHFATISDLQSQGTGGGILLRLHNMGRCKQGLMRE
 GEGHNVQEPARESELALQJGEEFVEEERELREFDEYTIYPTVEAHLISLIER
 VLCHRTLEGAWVGTOSAGCSKNSCFNPKFMFLRPESECVAVLQVPRRLQGL
 TRALGASPAVPLNGKDOAVGLHMKVEKRKISLPRVLSAPVAGTCAHAYDEIH
 LRCELSPGYVLAIVSTFLKDVPGGLHVRFGSKISLSTVRLATGASGTCAHAYDEIH
 ETVQJQGCRRAGQTGAGSRNFPSTYRCPNCLPVSVEGACPRITRTILOOCHRLSDSL
 HPIGHFVQPADGENDQCSLLDEPPLSCVPHRAQDEVSRLCLLSTVGNVRYIVSTY
 LPDTGETFTVTLATRIDQSHISOEMLGOLLOEVSMAYMKA"

CDS

BASE COUNT 537 a 713 c 749 g 529 t
 ORIGIN

alignment_scores:
 quality: 2969.00 Length: 671
 Ratio: 4.758 Gaps: 2
 Percent Similarity: 92.996 Percent Identity: 81.222

alignment_block:
 US-09-768-877-2 x BC010969

Align seg 1/1 to: BC010969 from: 1 to: 2528

```

1 MetarGAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
111 ATGCGGGGGGTCGGGGCCGAGACGCCGGGGAGCTCTCCGGGAGC 160
17 aAlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrProL 34
161 GGCATTCCTCCGCTCGACTCCTCGCTTTACACTTGTCCACGCTC 210
34 euAlaGlnPheArgGluAspIleThrThrPArgArgProGlnIleCys 50
211 TGGCCACATTCGGGAGAGACATCATTTGGAGAGACCCAGAAATCTGT 260
51 AlaThrProAlaGluPheProAspAspProArgGluGlnValLysG 67
261 GCCACACCTCAGCTGTTCCAGATACCATGGAGAGACAGGTGAACA 310
67 nGlyLeuLeuGlyAspCysTrpPheLeuCysAlaLysAlaLeuGlnL 84
311 AGGCTGCTGGAGATTCCTGCTGTCTGTGCTGTGCTGCTGCTGCTG 360
84 ySeraArgHisLeuLeuAspGlnValIleProProGlyGlnProSerTrp 100
361 AGAGTCACACACCTCCTGGACACAGCTTCCCTCCAGAGACAGCAG 410
101 AlaAspGlnGluTrpArgGlySerPheThrCysArgIleTrpGlnPheG 117
411 TCTGACCAAAATACCAAGCTTCTACCTGCTGATTTGGCATTTGG 460
117 yArgTrpAlaGluAlaThrThrAspAspArgLeuProCysLeuAlaGly 134
461 ACACGTGGAGAGAGACCATGATGATGATGATGATGATGATGATG 510
134 rGluCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
511 GACTCTGCTTTCCTCCGCTGACAGAGAGATGTTCTGCTGCTGCTG 560
151 LeuGluLysValIleValAlaLysValHisGlySerTrpGluHisLeuTrpAl 167
561 CTGGAAGAGGCTATGCTAAGCTGATGATGATGATGATGATGATG 610
167 aGlyGlnValAlaAspAlaLeuValAspLeuThrGlyGluLeuAlaGlu 184
611 AGGCAATGTCAGATGCTCTTATGATGATGATGATGATGATGATG 660
184 rGTrpAsnLeuLysGlyValAlaGlySerGlyGlnGlnAspArgPro 200
661 GGTGAGCTTGAAGAGATTAACGAAGCAGCGGCAGACAGACAGACC 710
201 GlyArgTrpGluHisArgThrCysArgGlnLeuLeuHisLeuLysAspG 217
711 AGTGTGGGAGACAGAACTTGTGCGACGATCTACCTCAGCAAGAG 760
217 nCysLeuIleSerCysCysValIleSerProArgAlaGlyAlaArgGlu 234
761 GTGCTATATCAGCTGCTGCTGCTTATGCCCAAGACAGAGGTCAG 810
234 euGlyGluPheHisAlaPheIleValSerAspLeuArgGluLeuGlnGly 250
811 TCGAGAGATTCATGCTTATCATCTCAGATTCGACAGAGCTAGAGGT 860
251 GlnAlaGlyGlnCysIleLeuLeuLeuArgIleGlnAsnProTrpGlyAr 267
861 CAGACTGCGCAGGATATCTCTGCTGCGATTCACAACCCCTGGGGCC 910
267 gArgCysTrpGlnGlyLeuTrpArgGluGlyGlnGlyTrpSerGln 284
911 GCGTTGTTGGCAGGCGCTCTGGAGAGAGAGAGAGAGAGAGAGAG 960
284 aAlaPalaAlaValAlaSerGluLeuLeuSerGlnLeuGlnGlyGlu 300
961 TAGAGCCAGTAAAGAGATCTAGCTGTGGCCCAACTCTCAGAGAGAG 1010

```

301 PheTyrValGluGluGluPheLeuArgGluPheAspGluLeuThrVa 317
 1011 TTCTGGTTCGAGGAAGAGGAGTTCTCTCAGGAGTTTGATGAGTCCACAT 1060
 317 lGlyTyrProValThrGluAlaGlyHisLeuGlnSerLeuTyrThrGluA 334
 1061 CGGCTACCCAGTCACAGAGCGCGGCCACCTACAGAGTCTCCACACAGAGA 1110
 334 rgLeuLeuCysHisThrArgAlaLeuProGlyAlaTyrValLysGlyGln 350
 1111 GGGTGTGTGCCATACCGCGGACACTGCCCTGGTGGCTGGGTGACAGGGCAG 1160
 351 SerAlaGlyGlyCysArgAsnAsnSerGlyPheProSerAsnProLysPh 367
 1161 TCAGCAGGAGGCTGCGGGAACAACAGTTCTTCCCTGCAACCCCAAGTT 1210
 367 enrPLeuArgValSerGluProSerGluValTyrIleAlaValLeuGlnA 384
 1211 CTGGTTACGGCTCTTGGGAACCCAGCGAGGTGTGTGGCTGTCTTCAGA 1260
 384 rgSerArgLeuHisAlaAlaAspTyrAlaGlyArgAlaArgAlaLeuVal 400
 1261 GACCCGGAGGCGCTTA.....GTGGCCAGACTCGGGCACTGGCG 1301
 401 GlyAspSerHisThrSerTyrProAlaSerIleProGlyLysHisTy 417
 1302 GGT.....GCCAGTCTCCACCGGTGAACCTCCAGGCAAGACTA 1342
 417 rGlnAlaValGlyLeuHisLeuTyrLysValGluLysArgArgValAsnL 434
 1343 CCAGGCTGTGGCCCTGCACATCTGGAAGGTAGAGAAACGGAAGATCAGCC 1392
 434 euProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAla 450
 1393 TGCCCCAGAGTCTGTGCACCCCTGTGGCTGGCACTGCATGCCATGCG 1442
 451 TyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTyrTyrLe 467
 1443 TATGATCGTGAGATCCACTTGGTGTGTGAGCTCTCACCAGGCTACTACT 1492
 467 uAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeuLeuA 484
 1493 GGCCGTCCCTAGCACCTTTTGAAGGATGTGCCAGGGCAGTTCTCTGCTCA 1542
 484 rgValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAla 500
 1543 GAGTCTTCTCCACTGGGAAATCTCCCTCAGTGGCGTCCAGCTGGCCACC 1592
 501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyGluTyrPglyThrVa 517
 1593 AAGGTTGCATCGCTGGAACAGCCCTGCCTGCAGCGAGTGGGAGACTGT 1642
 517 lGlnLeuArgGlySerTyrArgValGlyGlnThrAlaGlyGlySerArgA 534
 1643 GCAGTTGAGGGCTCTGGAGAGCTGGCCAGACAGCTGGGGGCGCAGGA 1692
 534 snPheAlaSerTyrProThrAsnProCysPheProPheSerValProGlu 550
 1693 ACTTTGGCTCTTACCCCTGCATCTCCCTGCCCTCTCTCTCTCTCTCTCT 1742
 551 GlyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArgPr 567
 1743 GGTGTCGCCCGCTACATCCGTATACCTACACCAACACTGCCGCT 1792
 567 oSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValProG 584
 1793 CAGTGACAGCAGCTGCAACCCCATTTGGTTTCCATGTCTTTCAGGTTCCAG 1842
 584 lucGlyGlyArgSerGlnAspAlaProProLeuLeuGlnGluProLeu 600
 1843 CAGACGGTGCAGAACGAGCGGTGTTCCCTGCTGCTCCAGGAGCCACTG 1892

601 LeuSerCysValProHisArgTyrAlaGlnGluValSerArgLeuCysLe 617
 1893 CTAAGCTGTGTACCATCGCTACGCCAGGAAGTGAGCCGCTCTGCCT 1942
 617 uLeuProAlaGlyThrTyrLysValValProSerThrTyrLeuProAspT 634
 1943 CCTTTCTGTGGGAAGTACAGAGATTGTTCCCTCCACCTACCTGCCAGATA 1992
 634 hrGluGlyAlaPheThrValThrIleAlaThrArgIleAspArgProSer 650
 1993 CAGAGGGTACCTTCACGGTAACCATAGCAACCAAGATCGATAGGCAGTCC 2042
 651 IleHisSerGlnGluMetLeuGlyGlnPheLeuGlnGluValSerValMe 667
 2043 ATCCACAGCCAGGAGATGCTGGCCAGCTGCTCCAGGAGGTCTCCTTTAT 2092
 667 tAlaValMetLys 671
 2093 GGCAGTGTATGAAA 2105

OM of: US-09-768-877-2 to: N.Geneseq.032802.* out_format : pfs

Date: Sep 17, 2002 6:39 AM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+g2n.model -DEV=xlp
-O=/cgn2.1/USPTO/spool/US09768877/runat.16092002.152812.8411/app-query.fasta.1.739
-DB=N.Geneseq.032802 -OFMT=fastp -SUFFIX=ring -GAPOP=12.000
-CAPEXT=4.000 -MINMATCH=0.100 -LOOPTOL=0.000 -LOOEXT=0.000
-GAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=bloms62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pcr
-THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTEXT=pfs
-USER=US09768877.ecgn1.1.690 -NCPU=6 -ICPU=3 -LONGLOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-768-877-2
Query length: 672
Database: N.Geneseq.032802.*
Database sequences: 1736436
Database length: 858457221
Search time (sec): 278.420000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	! Documentation
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AA27476 +		3644.00	3995.94	3.2e-214	26
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AA250930 +		3644.00	3990.54	6.3e-214	25
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAK94261 +		3668.00	3983.95	1.5e-213	24
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AA27481 +		3604.00	3947.11	1.7e-211	23
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AA27477 +		3290.00	3602.97	2.4e-192	22
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AA27480 +		3239.50	3546.72	3.3e-189	25
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AA27479 +		3090.50	3384.30	3.7e-180	22
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AA27484 +		2963.00	3243.18	2.7e-172	25
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:AA591171 +		2726.50	2887.82	4.5e-158	15
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA23883 +		2719.50	2977.09	1.8e-157	2
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA27478 +		2715.50	2973.45	2.8e-157	20
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA27475 +		2487.50	2725.47	1.8e-143	1
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AA27472 +		1989.00	2147.74	2.8e-111	49
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA57303 +		1251.00	1372.90	4.0e-68	8
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH9362 +		964.00	1059.15	1.2e-50	7
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH9362 +		879.00	968.88	1.3e-45	53
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AA575397 +		654.00	722.96	6.4e-32	4
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA57188 +		660.50	727.77	3.4e-32	61
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA57305 +		644.50	699.51	1.3e-30	2
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA57304 +		644.50	699.31	1.3e-30	2
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:AA591170 +		644.50	699.31	1.3e-30	21
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA57303 +		644.50	699.05	1.4e-30	2
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA57302 +		644.50	698.86	1.4e-30	2
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA52158 +		644.50	698.40	1.5e-30	2
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA525803 +		644.50	698.40	1.5e-30	2
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AA274910 +		644.50	698.40	1.5e-30	23
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA57306 +		644.50	696.80	1.8e-30	2
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA509261 +		641.50	694.41	2.5e-30	2
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA53142 +		641.50	694.41	2.5e-30	2
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAV04202 +		635.00	688.88	5.0e-30	21
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAV04203 +		635.00	688.77	5.1e-30	21
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:AA57167 +		631.00	684.48	8.9e-30	24
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:AA53844 +		631.00	683.16	1.1e-29	24
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AA57383 +		618.00	666.90	8.5e-29	3
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV5491 +		617.00	670.79	5.1e-29	17
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV5491 +		617.00	670.79	5.1e-29	17
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV5490 +		617.00	667.77	7.6e-29	24
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV44010 +		617.00	667.77	7.6e-29	24
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AA210396 +		615.50	666.51	8.9e-29	23

/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA560948 + 615.00 658.61 2.4e-
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA541201 + 614.00 680.13 1.6e-
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV44013 + 608.00 657.62 2.8e-21
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH09519 - 605.50 668.17 7.2e-
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:AA592630 - 601.50 648.38 9.1e-
seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AA27476

seq_documentation_block:

ID AAA27476 standard; cDNA; 2620 BP.

AAA27476;

15-AUG-2000 (first entry)

Human calpain 10a cDNA.

NIDDM1; non-insulin-dependent diabetes mellitus; CAPN10 gene;

calpain 10; calpain 10a; diapiain-1; cysteine protease;

chromosome 2; human; diagnosis; therapy; ss.

Homo sapiens.

Key Location/Qualifiers

CDS 178..2196

FT /*tag= a

MO200023603-A2.

27-APR-2000.

21-OCT-1999; 99WO-US24890.

21-OCT-1998; 98US-0105052.

13-MAY-1999; 99US-0134175.

(ARCH-) ARCH DEV CORP.

Polonsky KS, Horikawa Y, Oda N, Sreeman S, Zhou Y, Otsani K;

Hants Cl, Bell G;

WPI; 2000-339702/29.

P-PSDB; AA79567.

Method for screening for type 2 diabetes mellitus comprises detecting a

polymorphism in a calpain encoding nucleic acid segment or a

protease-encoding nucleic acid segment -

Claim 65; Page 219-220; 257pp; English.

The present sequence is that of cDNA corresponding to the major

transcript of the human calpain 10 gene, CAPN10 (see AA27475),

excluding the polyA tract. The sequence contains an open reading

frame that encodes a protein of 672 amino acids (see AA79567),

designated calpain 10a. The CAPN10 gene consists of 15 exons

spanning 32 kb. A complex pattern of alternative splicing

generates proteins of 672, 544, 517, 513, 444, 274, 139 and 138

amino acids (see AA79567-74), designated calpain 10a to 10h,

respectively. Isoform 10a is encoded by exons 1-7 and 9-13 of

the CAPN10 gene. Transcripts encoding calpain 10a are the most

abundant in the various tissues examined. The invention concerns

the identification of genes responsible for type 2 diabetes for use

in diagnostic and therapeutic applications. A G-to-A polymorphism

in intron 3 of the CAPN10 gene in the NIDDM region of chromosome 2

(UCSNP-43) shows evidence for linkage to type 2 diabetes. Claimed

methods for screening for a propensity for type 2 diabetes mellitus

are based on detection of a polymorphism in a calpain encoding

nucleic acid, especially UCSNP-10 of the CAPN10 gene. Methods are

also claimed for: producing calpain 10 polypeptides using calpain

10-encoding polynucleotides, including the present sequence;

identifying modulators of calpain activity using calpain 10

polypeptides and nucleic acids; and using these modulators to treat

diabetes, in particular through the regulation of an insulin

CC secretory response or insulin mediated glucose transport.
XX
SQ Sequence 2620 BP; 448 A; 819 C; 899 G; 454 T; 0 other;

alignment_scores:
Quality: 3649.00 Length: 672
Ratio: 5.430 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-768-877-2 x AAA27476 ..

Align seg 1/1 to: AAA27476 from: 1 to: 2620

1 MetArgAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
178 ATCGGG 227
17 alaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrProL 34
228 CGCCTTCCCGCGCGGGGACTCTCGCTCTTCTGCGACTTGTCTACGCCG 277
34 euAlaGlnPheArgGluAspIleThrTrpArgArgProGlnGluIleCys 50
278 TGCCCCAGTTCCCGGAGGACATCACGTGGAGCGGGCCCGGAGATTGT 327
51 AlaThrProArgLeuPheProAspAspProArgGluGlyGlnVallysl 67
328 GCCACACCCCGGCTGTTCCAGATGACCACCGGAAGGCGAGGTGAAGCA 377
67 nGlyLeuLeuGlyAspCysTrpPheLeuCysAlaCysAlaAlaLeuGlnL 84
378 GGGGGTGTGGGGGATGTTGTTCTGTTGCTGCTGCGCGCGCTGCACA 427
84 ySerArgHisLeuLeuAspGlnValIleProProGlyGlnProSerTrp 100
428 AGACAGGCGACCTCTCTGGACACAGGTCTTCTCCGGGACAGCGAGCTG 477
101 AlaAspGlnGluTyArgGlySerPheThrCysArgIleTrpGlnPheG 117
478 GCGGACACAGGAGTACCGGGGCTCTTACCTGTGCGATTGGCAGTTTGG 527
117 yArgTrpValGluValThrThrAspAspArgLeuProCysLeuAlaGly 134
528 AGCTGGGTGGAGGTGACACAGATGACCGCTGCGCTGCGCTTGCAGGCA 577
134 rgLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
578 GACTCTGTTCTCCCGCTGCCAGAGGAGGATGTTCTGGCTCCCTTTA 627
151 LeuGluLysValTyAlaLysValHisGlySerTyArgGluHisLeuTrpAl 167
628 CTGGAAAGGTTCTACGCCAAGGTCTTACGGTCTTACGAGCACCTGTGG 677
167 aGlyGlnValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGlu 184
678 CGGGCAGGTGGGGAGTGGCTGGTGGACTGACCGCGGGCGCTGGCAGAA 727
184 rgTrpAsnLeuLysGlyValAlaGlySerGlyGlnGlnAspArgPro 200
728 GATGAACTGAAGCGGTAGCAGGAGCGGAGGCGGAGGAGGAGGAGGAG 777
201 GlyArgTrpGluHisArgThrCysArgGlnLeuLeuHisLeuLysAspG 217
778 GCGCGCTGGGAGCACAGGACTTGTGGCAGCTGTCTCCACCTGAAGGACA 827
217 nCysLeuIleSerCysValLeuSerProArgAlaGlyAlaArgGluL 234
828 GTGTTGATCAGCTGCGTCTCAGCCCCAGAGCAGGTGCGCGGAGC 877
234 euGlyGluPheHisAlaPheIleValSerAspLeuArgGluLeuGlnGly 250
|||||

878 TGGGGGAGTTCCATGCTTCAATTGCTCGGACCTCGGGAGCTCCAGGGT 927
251 GlnAlaGlyGlnCysIleLeuLeuLeuArgIleGlnAsnProTrpGlyAr 267
928 CAGCGGGGCGAGTGCATCTGCTGCTGCGGATCCAGAACCCTCGGGCGC 977
267 gArgCysTrpGlnGlyLeuTrpArgGluGlyGlyGlyTrpSerGlnV 284
978 GCGGTCTGCGAGGGGCTCTGGAGAGAGGGGGGTGAAGGTGGAGCAGG 1027
284 alaSpAlaAlaValAlaSerGluLeuLeuSerGlnLeuGlnGlyGlu 300
1028 TAGATCAGCGGTAGCATCTGAGCTCTGTCAGCTCCAGGAAGGGAG 1077
301 PheTrpValGluGluGluPheLeuArgGluPheAspGluLeuThrVa 317
1078 TTTCTGGGTGGAGGAGGAGGAGTTCTTCAGGGAGTTTGACGAGCTCACCG 1127
317 lGlyTyProValThrGluAlaGlyHisLeuGlnSerLeuTyThrGluA 334
1128 TGGCTACCGGGTCACGAGGCGCGCCACTGACAGACCTCTACACAGAGA 1177
334 rgLeuLeuCysHisThrArgAlaLeuProGlyAlaTrpValLysGlyGln 350
1178 GGCCTGCTCTGCCATACGCGGGCGCTGCTGGGCGCTGGGTCAAGGGCCAG 1227
351 SerAlaGlyGlyCysArgAsnAsnSerGlyPheProSerAsnProLysph 367
1228 TCAGCAGGAGGCTGCCGGGAACACAGCGCTTTCCACAGCAACCCCAATT 1277
367 eTrpLeuArgValSerGluProSerGluValTyIleAlaValLeuGlnA 384
1278 CTGGCTGGGGTCTCAGAACCGAGTCAGGTGTACATTGCGCTCTGCAGA 1327
384 rgSerArgLeuHisAlaAlaAspTrpAlaGlyArgAlaAlaVal 400
1328 GATCCAGGCTGCACGCGCGACTGGGAGCGCGCGCGCGGCTGCTGCTG 1377
401 GlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHisTy 417
1378 GGTGACAGTCATCTCTGTTGGAGCGCGGAGCATCCCGGGCAAGCACTA 1427
417 rGlnAlaValGlyLeuHisLeuTrpLysValGluLysArgValAsnL 434
1428 CCAGGCTGTGGGTCTGCACCTCTGGAAGGTACAGAAGCGCGCGGTCAATC 1477
434 euProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAla 450
1478 TGCCCTAGGGTCTGTCCATGCCCGCGCTGGTGGCACCGCGTGCATGCA 1527
451 TyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTyTrile 467
1528 TACGACCGGGAGTCCACCTGCGTTGTGAGCTCTCACCGGGGTACTTACCT 1577
467 uAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeuLeu 484
1578 GGCTGCCCGAGCACCTTCTGAAGGAGCGCGCAGGGAGTTCTCTGCTCC 1627
484 rgValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAla 500
1628 GAGTCTTCTTACCGGGGAGTCTCCCTTAGCGCATCAGGGCAGTGCGCC 1677
501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyGluTrpGlyThrVa 517
1678 AAGAACACACCCCGGGGAGCTGCTGCGGGGAGTGGGGGAGCGT 1727
517 lGlnLeuArgGlySerTrpArgValGlyGlnThrAlaGlyGlySerArg 534
1728 GCAGCTACGGGGTCTTGGAGAGTGGGCGCAGACGGCGGGGGGCGACAGGA 1777
534 snPheAlaSerTyProThrAsnProCysPheProPheSerValProGlu 550
1778 ACTTTGCTCATACCCCAACCAACCCCTGCTTCCCTTCTCGGTCGCCGAG 1827

134 rgLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
|||||
549 GACTCTGTTTCCCGCTGCGACAGGAGGATGTCTTGCTCCCTTA 598
151 LeuGluLysValTyrAlaLysValHisGlySerTyrGluHisLeuTrpAl 167
|||||
599 CTGAAAGGCTTACGCCAAGGTCCATGGGTCTACGAGCACCCTGTGGGC 648
167 aglyGlnValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGluA 184
|||||
649 CGGGCAGGTGGCGATGCCCTGTGTGACCTGACCGCGGCGCTGGCAGAA 698
184 rgTrpAsnLeuLysGlyValAlaGlySerGlyGlnGlnAspArgPro 200
699 GATGGAACCTGAAGGCGGTAGCAGGAAGCGGAGCGGCGCAGCAGGCGG 748
201 GlyArgTrpGluHisArgThrCysArgGlnLeuLeuHisLeuLysAspG 217
|||||
749 GGCCGCTGGGAGCACAGGACTTGTTCGCCAGCTCTCCACCTGAAGGACCA 798
217 nCysLeuIleSerCysCysValLeuSerProArgAlaGlyAlaArgGluL 234
|||||
799 GTGTCTGATCAGCTGTGGTCTGAGCCAGCCAGAGCAGGTGCCGGGAGC 848
234 euGlyGluPheHisAlaPheIleValSerAspLeuArgGluLeuGlnGly 250
849 TGGGGAGTTCATGCTCTTCATTGTCTCGACCTGCGGAGCTCCAGGCT 898
251 GlnAlaGlyGlnCystIleLeuLeuArgIleGlnAsnProTrpGlyAr 267
|||||
899 CAGCGGGCGAGTGCATCTGCTGTGCGGATCCAGAACCCCTGGGGCGG 948
267 gArgCysTrpGlnGlyLeuTrpArgGluGlyGlyGlyTrpSerGlnV 284
|||||
949 GCGGTCTGGCAGGGGCTCTGGAGAGAGGGGTGAAGGTGGAGCCAGG 998
284 alaAspAlaValAlaSerGluLeuLeuSerGlnLeuGlnGlyGlu 300
999 TAGATCAGCGGTAGCATCTGAGCTCTCTCCAGCTCCAGGAAGGGGAG 1048
301 PheTrpValGluGluGluPheLeuArgGluPheAspGluLeuThrVa 317
|||||
1049 TTTCTGGTGGAGAGAGAGGTTCCTCAGGGAGTTTGACGAGCTCACCGT 1098
317 lGlyTrpValThrGluAlaGlyHisLeuGlnSerLeuTyrThrGluA 334
|||||
1099 TGCTACCCGGTCACGGAGCGCGCCACCTGCAGAGCCTCTACACAGAGA 1148
334 rgLeuLeuCysHisThrArgAlaLeuProGlyAlaTrpValLysGlyGln 350
1149 GGTGCTCTGCCATACGCGGGGCTGCTTGGGGCTGGGTCAAGGGCCAG 1198
351 SerAlaGlyGlyCysArgAsnAsnSerGlyPheProSerAsnProLysPh 367
|||||
1199 TCAGCAGGAGGCTGCCGGAAACACAGCGGCTTTCACAGCAACCCCAANT 1248
367 eTrpLeuArgValSerGluProSerGluValTyrIleAlaValLeuGlnA 384
|||||
1249 CTGGCTGCGGGTCTCAGAACCCAGGTGAGGTGTACATTCGCGCTCTGCAGA 1298
384 rgSerArgLeuHisAlaAlaAspTrpAlaGlyArgAlaArgAlaLeuVal 400
1299 GATCCAGGTGACGCGGGGACTGGGCAGCGCGGCGGCGGCGGCGGCTGG 1348
401 GlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHisTy 417
|||||
1349 GGTGACAGTCATCTCTGTGGAGGCCAGCGAGCATCCCGGGCAAGCACTA 1398
417 rGlnAlaValGlyLeuHisLeuTrpLysValGluLysArgArgValAsnL 434
|||||
1399 CCAGGCTGTGGTCTGCACCTCTGGAGGTAGAGAGCGCGGGGTCAATC 1448

434 euProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAla 450
|||||
1449 TGCTAGGTCTCTGTCCATGCCCGCGTGGTGGCACCGGTGCCATGCA 1498
451 TyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTyrTrpLe 467
|||||
1499 TAGCACCGGAGTCCACCTGCGTGTGAGCTCTCACCGGGCTACTACCT 1548
467 uAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeuLeuA 484
|||||
1549 GGTGTGCCCGACACCTTCTTGAAGACGCGCGGAGGTTCCTGCTCTCC 1598
484 rgValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAla 500
|||||
1599 GAGTCTTCTTACCGGGCGAGTCTCCCTTAGCGCCATCAGGCGAGTGGCC 1648
501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyGluTrpGlyThrVa 517
|||||
1649 AAGACACCGCGCGCGGCGAGCCCTGCTGCGGGGAGTGGGGGACCGT 1698
517 lGlnLeuArgGlySerTrpArgValGlyGlnThrAlaGlyGlySerArgA 534
|||||
1699 GCAGCTACGGGGTTCCTTGAGAGTGGCCAGACGCGGGGGGCGCAGCAGA 1748
534 snPheAlaSerTyrProThrAsnProCysPheProPheSerValProGlu 550
|||||
1749 ACTTGGCTCATACCCACACCCCTGCTTCCCTTCTCGGTCCCGAG 1798
551 GlyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArgPr 567
|||||
1799 GSCCCTGGCGCGCTGCTGCGCATCACTCTGCATCAGCACTGCGGGCC 1848
567 oSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValProG 584
|||||
1849 CAGTGCACCGAGTTCACCCCATCGCTTCCATATCTTCCAGGTCCAG 1898
584 luGlyClyArgSerGlnAspAlaProProLeuLeuLeuGlnGluProLeu 600
1899 AGGGTGAAGAGCAGGAGCAGCAGCCCACTGCTGCTGCAGGAGCGCTG 1948
601 LeuSerCysValProHisArgTyrAlaGlnGluValSerArgLeuCysLe 617
1949 CTGAGCTGCGTCCACATCGCTACGCCCGCAGGAGTGGCGGCTCTGCT 1998
617 uLeuProAlaGlyThrTyrLysValValProSerThrTyrLeuProAspT 634
1999 CCTGCTGGGGCACCTACAGGTGTGCGCTTCCACTACCTTGGCGGACA 2048
634 hrGluGlyAlaPheThrValThrIleAlaThrArgIleAspArgProSer 650
2049 CAGAGGGGCGCTTTCACAGTGCACCATCGCAACCCAGGATTGACAGGCCATCC 2098
651 lleHisSerGlnGluMetLeuGlyGlnPheLeuGlnGluValSerValMe 667
2099 ATTCACAGCCAGAGATGCTGGCGCATCTCTCCAAGAGGTCTCGCTCAT 2148
667 tAlaValMetLysThr 672
2149 GCAGTGTGATGAACCC 2164

seq_name: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AAK94261
seq_documentation_block:
ID AAK94261 standard; cDNA: 2596 BP.
XX
AC AAK94261;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human full-length cDNA, SEQ ID NO: 2880.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX

401 GlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHisTy 417
1348 GGTGACAGTCATACCTTCGTGGAGCCAGCAGCATCCCGGCAAGCACTA 1397
417 rGlnAlaValGlyLeuHisLeuTrpLysValGluLysArgValAsnL 434
1398 CCAGGCTGGGGTCTGCACCTCTGGAGGTAGAGAAGCGCGGGTCAATC 1447
434 euProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAla 450
1448 TGCTAGGCTCTGTCTCATGCCGCCCGCTGGTGGCACCGCTGCCATGCA 1497
451 TyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTyrTrpLe 467
1498 TAGCAGCGGAGTCCACCTGGCTGTGTGAGCTCTCACCGGGCTACTACCT 1547
467 uAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeuLeuA 484
1548 GGTGTCTCCACAGCACCCTCTGAAGGACGCGCCAGGGGAGTTCTCTGCTCC 1597
484 rgValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAla 500
1598 GAGTCTCTCTACCGGGCGAGTCTCCCTTAGCGCCATCAGGCGAGTGGCC 1647
501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyGluTrpGlyThrVa 517
1648 AAGAACACCGCCCGCGGCGAGCCCTGCTGCGGGGAGTGGGGACCGT 1697
517 lGlnLeuArgGlySerTrpArgValGlyGlnThrAlaGlyGlySerArgA 534
1698 GCAGCTACCGGGTCTTGTGAGAGTGGCCAGACGCGGGGCGCAGCAGGA 1747
534 snPheAlaSerTyrProThrAsnProCysPheProPheSerValProGlu 550
1748 ACTTTGCTCATACCCACCCACCCCTGCTTCCCTCTCGGTCGCCGAG 1797
551 GlyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArgPr 567
1798 GGCCCTGGCCCGCGTGTGCTGCGCATCACTTGCATCAGCAGCTGCGGGC 1847
567 oSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValProG 584
1848 CAGTGACACCGAGTTCACCCCATCCGCTTCCATATCTTCCAGGTCCGAG 1897
584 luGlyGlyArgSerGlnAspAlaProProLeuLeuGlnGluProLeu 600
1898 AGGTTGGAGGACGACGAGGACGACCCCACTCTCTGCTGAGGAGCGCTG 1947
601 LeuSerCysValProHisArgTyrAlaGlnGluValSerArgLeuCysLe 617
1948 CTGAGCTGGTGCACATCGCTACGCCAGGAGGTGAGCGGCTCTGCT 1997
617 uLeuProAlaGlyThrTyrLysValValProSerThrTyrLeuProAspT 634
1998 CCGTCTGGGGACCTACAGGTTGTGCTTCCACCTACCTTCCCGGACA 2047
634 hrGluGlyAlaPheThrValThrIleAlaThrArgIleAspArgProSer 650
2048 CAGAGGGGCTTACAGTGACCATCGACACCCAGGATGACAGGCGCATCC 2097
651 IleHisSerGlnGluMetLeuGlyGlnPheLeuGlnGluValSerValMe 667
2098 ATTCACAGCCAGGAGATGCTGGGCGAGTCTCTCCAAAGAGGTCTCGCTCAT 2147
667 tAlaValMetLysThr 672
2148 GCAGTGATGAANACC 2163

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAA27481
seq_documentation_block:
ID AAA27481 standard; cDNA; 2455 BP.

XX AAA27481;
XX AC
XX DT 15-AUG-2000 (first entry)
XX DE Human calpain 10f cDNA.
XX KW NIDDM1; non-insulin-dependent diabetes mellitus; CAPN10 gene;
XX KW calpain 10; calpain 10f; diapain-1; cysteine protease;
XX KW chromosome 2; human; diagnosis; therapy; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT CDS 42..866
XX FT /*tag= a
XX PN WO200023603-A2.
XX PD 27-APR-2000.
XX PF 21-OCT-1999; 99WO-US24890.
XX PR 21-OCT-1998; 98US-0105052.
XX PR 13-MAY-1999; 99US-0134175.
XX PA (ARCH-) ARCH DEV CORP.
XX PI Polonsky KS, Horikawa Y, Oda N, Sreenan S, Zhou Y, Otani K;
XX PI Hanis CL, Bell GI;
XX DR WPI; 2000-339702/29.
XX DR P-PSDB; AAY79572.
XX PT Method for screening for type 2 diabetes mellitus comprises detecting a
XX PT polymorphism in a calpain encoding nucleic acid segment or a
XX PS protease-encoding nucleic acid segment -
XX PS Claim 65; Page 232-232; 257pp; English.
XX CC The present sequence is that of cDNA corresponding to a minor
XX CC transcript of the human calpain 10 gene, CAPN10 (see AA227475).
XX CC The sequence contains an open reading frame that encodes a protein
XX CC of 274 amino acids (see AAY79572), designated calpain 10f. The
XX CC CAPN10 gene consists of 15 exons spanning 32 kb. A complex pattern
XX CC of alternative splicing generates proteins of 672, 544, 517, 513,
XX CC 444, 274, 139 and 138 amino acids (see AAY79567-74), designated
XX CC calpain 10a to 10h, respectively. Calpain 10f is encoded by exons
XX CC 1-3*, 4-7 and 9-13 of the gene. The invention concerns the
XX CC identification of genes responsible for type 2 diabetes for use in
XX CC diagnostic and therapeutic applications. A G-to-A polymorphism in
XX CC intron 3 of the CAPN10 gene in the NIDDM1 region of chromosome 2
XX CC (UCSNP-43) shows evidence for linkage to type 2 diabetes. Claimed
XX CC methods for screening for a propensity for type 2 diabetes are
XX CC based on detection of a polymorphism in a calpain encoding nucleic
XX CC acid, especially UCSNP-10 of the CAPN10 gene. Methods are also
XX CC claimed for: producing calpain 10 polypeptides using calpain
XX CC 10-encoding polynucleotides, including the present sequence;
XX CC identifying modulators of calpain activity using calpain 10
XX CC polypeptides and nucleic acids; and using these modulators to
XX CC treat diabetes, in particular through the regulation of an insulin
XX CC secretory response or insulin mediated glucose transport.
XX SO Sequence 2455 BP; 429 A; 769 C; 827 G; 430 T; 0 other;

alignment_scores:
Quality: 3604.00 Length: 673
Ratio: 5.395 Gaps: 2
Percent Similarity: 99.257 Percent Identity: 99.257
alignment_block:
US-09-768-877-2 x AAA27481

Align seq 1/1 to: AAA27481 from: 1 to: 2455

```

1 MetArgAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgSpAl 17
42 ATCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 91
17 AlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrPro 34
92 CGCCTTCCCGCGCGGACCTCCCTCTCTCTGCACTTGTCAGCCCGC 141
34 euAlaGlnPheArgGluAspIleThrTrpArgArgProGlnIleCys 50
142 TGCCCAAGTTCGCGAGACATCAGTGAAGGGGCCCGCAGGAGATTGT 191
51 AlaThrProArgLeuPheProAspProArgGluGlnValIleLysG 67
192 GCACACACCCGGGCTTTCAGATGACCCACGGGAAGGCGAGGTGAAGA 241
67 nGlyLeuLeuGlyAspCysTrpPheLeuCysAlaAlaLeuGlnL 84
242 GGGGCTGGCGGGGATGCTGCTCTGCTGCTGCTGCTGCTGCTGCTG 291
84 ySSerArgHisLeuLeuAspGlnValIleProProGlyGlnProSer 100
292 AGACGACGACCTCTCGACACAGTCATTCCTCGGACAGCCGAGCTGG 341
101 AlaAspGlnGluTrpArgGlySerPheThrCysArgIleTrpGlnPhe 117
342 GCGGACACAGAGTACCGGGGCTCCTTCACCTCTGCACTTGCACTTGG 391
117 yArgTrpAlaGluValIleThrAspAspArgLeuProCysLeuAlaG 134
392 ACGGTGGGGAGGTGACACAGATGACCGCTGCGCTGCTGCTGCTGCA 441
134 rGLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuPro 150
442 GACTGTGTTTCCTCCGCTGCGACAGAGGAGATGTTCTGGCTCCCTTA 491
151 LeuGlnLysValTyrAlaLysValHisGlySerTyrGlnHisLeuTP 167
492 CTGAAA.....AAGGTCATGGGCTGACAGACACCTGTGG 529
167 LagGlnValAlaAspAlaLeuValAspLeuThrGlyLeuAlaGlu 183
530 CCGGGACAGTGGAGATCCCTGTGGACTGACCGGGCTGCTGCAAGA 579
184 ArgTrpAsnLeuLysGlyValAlaGlySerGlyGlnGlnAspArgPr 200
580 AGATGGAACTGMAAGGGCTGACAGAGAGGAGGCGACAGCAAGGCC 629
200 OGlyArgTrpGluHisArgThrCysArgGlnLeuLeuHisLeuLysAs 217
630 AGCGCGCTGGAGACAGAGACTTTCGCGACCTGCTCACCCTGAAGACC 679
217 LncysLeuLeuSerCysCysValLeuSerProAlaGlyAlaArgGlu 233
680 AGTCTTGATAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 729
234 LeuGlyGluPheHisAlaPheIleValSerAspLeuArgGluLeuGln 250
730 CTGGGGAGATTCATGCTTCATGTTCTCGGACTGCGGAGCTCCAGGG 779
250 yGlnAlaGlyLncysIleLeuLeuLeuArgIleGlnAsnProTrpGly 267
780 TCAGGCGGGCAGTGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 829
267 rGArgCysTrpGlnGlyLeuTrpArgGluGlyGlyGlyLysTrpSerGln 283
830 GGGGCTGCTGGAGGGCTCTGAGAGAGGGGGTGAAGGGTGAAGCCAG 879
284 ValAspAlaAlaValAlaSerGluLeuLeuSerGlnLeuGlnGlyG 300
880 GTAGATCAGCGGTAGCATCTGAGCTCTGTCCACTCCAGAGAGGA 929
300 uPheTrpValGluGluGluGluLeuPheLeuArgGluPheAspGluLeu 317
930 GTTCTGGGTGGAGAGAGAGAGATTCCTCAAGGGATTGAAGAGCTACCG 979
317 aIGlyTrpProValIleThrGluAlaGlyHisLeuGlnSerLeuTyrThr 333
980 TTGGCTACCGGGTCAAGAGGCGGCGCCTGACAGGCTTCAACAGAGAG 1029
334 ArgLeuLeuCysHisThrArgAlaLeuProGlyAlaIleTrpValLys 350
1030 AGCGTCTCTGCCATACGGGGGCGCTGCTGGGGCTGGGGTCAAGGGCCA 1079
350 nSerAlaGlyGlyCysArgAsnAsnSerGlyPheProSerAsnProLys 367
1080 GTACAGCAGAGGCTCGCGGACACAGACAGCGGCTTCCACAGCAACCC 1129
367 heTrpLeuArgValSerGluProSerGluValTyrIleAlaValLeuGln 383
1130 TCTGGCTGGGGTCTGAGAACAGAGAGGTGTACATTCCCTCTGACAG 1179
384 ArgSerArgLeuHisAlaAlaAspTrpAlaGlyArgAlaArgAlaLeu 400
1180 ACATCGAGGCTGCACGCGGGGAGCTGGCGACGCGGGCCCGGCACTGGT 1229
400 LglYAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHis 417
1230 GGGTACAGTCACTCTGCGAGCCAGCAGCAGATCCCGGGCAAGCACT 1279
417 yGlnAlaValGlyLeuHisLeuTyrLysValGlnLysArgArgValAsn 433
1280 ACCAGGCTGGGTGTGCACTCTGGAAGTGAAGAAGGGGGGTCAAT 1329
434 LeuProArgValLeuSerMetProProValAlaGlyThrAlaCysHis 450
1330 CTGCCTAAGGTCTCTCCATGCCCCCGTGGCTGGACCGCTGCCATGC 1379
450 aTyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTyrYL 467
1380 ATACAGCGGGAGGTCACTGCTGTTGTAGCTCTACCGGGCTACTACC 1429
467 euAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeu 483
1430 TGGCTGTCCCAAGCACTTCTGAAAGAGCGCCAGGGAGATTCTGTGCT 1479
484 ArgValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaVal 500
1480 CGAGCTTCTTACCGGGCGAGTCTTCCCTTAGCGGCATCAGGCGAGTGG 1529
500 aLysAsnThrThrProGlyAlaAlaLeuProAlaGlyLysTrpGlyThr 517
1530 CAAGAACAACACCCCGGGGCAAGCTCTGCGGGGAGTGGGGAGACCG 1579
517 aGlnLeuArgGlySerTrpArgValGlyGlnThrAlaGlySerArg 533
1580 TGCAGCTACGGGGTCTTGGAGAGTGGCCAGAGCGGGGGGAGAGCAG 1629
534 AsnPheAlaSerTyrProThrAsnProCysPheProPheSerValPro 550
1630 AACTTGGCTCATACCCACCAACCCCGTCTCCCTTCTGGTCCCGGA 1679
550 uGlyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArg 567
1680 GGGGCTGGGGCGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCT 1729
567 roSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValPro 583
1730 CCACTGACACCGAGTTCACCCCAATGGGCTTCCATCTCCAGGTCCA 1779
584 GluGlyLysArgSerGlnAspAlaProProLeuLeuLeuGlnGluPro 600
1780 GAGGTGAAGAGACCAAGACGACCCCTGCTGCTGCAAGAGCGCT 1829

```

600 uLeuSerCysValProHisArgTyrAlaGlnGluValSerArgLeuCysL 617
|||||
1830 GCTGAGCTGGTCCACATCGCTACGCCAGAGGTGAGCGGCTCTGCC 1879
617 euLeuProAlaGlyThrTyrLysValValProSerThrTyrLeuProAsp 633
|||||
1880 TCCTGCTGCTGAGGACCTACAGGTTGTGCGCTCCACCTACCTGCGGAC 1929
634 ThrGluValAlaPheThrValThrIleAlaThrArgIleAspArgProSe 650
|||||
1930 ACAGAGGGGCGCTTCACAGTGACCATCGCAACAGGATTGACAGGCCATC 1979
650 rIleHisSerGlnGluMetLeuGlyGlnPheLeuGlnGluValSerValM 667
|||||
1980 CATTACAGCCAGAGATGCTGGCGCAGTTCCTCCAGAGGTTCTCGTCA 2029
667 etalaValMetLysThr 672
|||||
2030 TGGCAGTGATGAAACC 2046

seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:AAA27477

seq_documentation_block:
ID AAA27477 standard; cDNA; 2297 BP.

XX AC AAA27477;
XX

XX DT 15-AUG-2000 (first entry)
XX

XX DE Human calpain 10b cDNA.
XX

XX KW NIDDM1; non-insulin-dependent diabetes mellitus; CAPN10 gene;
XX

XX KW calpain 10; calpain 10b; diapain-1; cysteine protease;
XX

XX KW chromosome 2; human; diagnosis; therapy; ss.
XX

XX OS Homo sapiens.
XX

XX FH Key Location/Qualifiers
XX

XX FT CDS 42..1676
XX

XX FT /*tag= a
XX

XX PN WO200023603-A2.
XX

XX PD 27-APR-2000.
XX

XX PF 21-OCT-1999; 99WO-US24890.
XX

XX PR 21-OCT-1998; 98US-0105052.
XX

XX PR 13-MAY-1999; 99US-0134175.
XX

XX PA (ARCH-) ARCH DEV CORP.
XX

XX PI Polonsky KS, Horikawa Y, Oda N, Sreenan S, Zhou Y, Otani K;
XX

XX PI Hanis CL, Bell GI;
XX

XX DR WPI; 2000-339702/29.
XX

XX DR P-PSDB; AAY79568.
XX

XX PT Method for screening for type 2 diabetes mellitus comprises detecting a
XX

XX PT polymorphism in a calpain encoding nucleic acid segment or a
XX

XX PT protease-encoding nucleic acid segment -
XX

XX PS Claim 65; Page 222; 257pp; English.
XX

XX CC The present sequence is that of cDNA corresponding to a
XX

XX CC transcript of the human calpain 10 gene, CAPN10 (see AA227475).
XX

XX CC The sequence contains an open reading frame that encodes a protein
XX

XX CC of 544 amino acids (see AAY79568), designated calpain 10b. The
XX

XX CC CAPN10 gene consists of 15 exons spanning 32 kb. A complex pattern
XX

XX CC of alternative splicing generates proteins of 672, 544, 517, 513,
XX

XX CC 444, 274, 139 and 138 amino acids (see AAY79567-74), designated
XX

XX CC calpain 10a to 10h, respectively. Calpain 10b is encoded by exons

CC 1-7, 9, 10* and 11-13 of the gene. It is readily detectable in
CC many tissues, including skeletal muscle and islets. The invention
CC concerns the identification of genes responsible for type 2 diabetes
CC for use in diagnostic and therapeutic applications. A G-to-A
CC polymorphism in intron 3 of the CAPN10 gene in the NIDDM1 region of
CC chromosome 2 (UCSNP-43) shows evidence for linkage to type 2
CC diabetes. Claimed methods for screening for a propensity for type
CC 2 diabetes are based on detection of a polymorphism in a calpain
CC encoding nucleic acid, especially UCSNP-10 of the CAPN10 gene.
CC Methods are also claimed for: producing calpain 10 polypeptides
CC using calpain 10-encoding polynucleotides, including the present
CC sequence; identifying modulators of calpain activity using calpain
CC 10 polypeptides and nucleic acids; and using these modulators to
CC treat diabetes, in particular through the regulation of an insulin
CC secretory response or insulin mediated glucose transport.
XX
SQ Sequence 2297 BP; 404 A; 713 C; 772 G; 408 T; 0 other;

alignment_scores:
Quality: 3290.00 Length: 672
Ratio: 5.350 Gaps: 1

Percent Similarity: 91.518 Percent Identity: 91.518

alignment_block:
US-09-768-877-2 x AAA27477

Align seg 1/1 to: AAA27477 from: 1 to: 2297

1 MetArgAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
|||||
42 ATGCGGCG 91
17 alaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrProL 34
|||||
92 CGCCTTCCCG 141
34 euAlaGlnPheArgGluAspIleThrTrpArgArgProGlnGluIleCys 50
|||||
142 TGCCCGAGTTCGCGGAGGACATCAGTGGAGCGCGCGCGCGCGCGCGCGCG 191
51 AlaThrProArgLeuPheProAspAspProArgGluGlyGlnValLysG 67
|||||
192 GCCACACCCCGCGCTGTTTCCAGATGACCCAGCGGAGGCGCGCGCGCG 241
67 nGlyLeuLeuGlyAspCysTrpPheLeuCysAlaCysAlaAlaLeuGlnL 84
|||||
242 GGGGCTGCTGGGGGATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 291
84 ySerArgHisLeuLeuAspGlnValIleProProGlyGlnProSerTrp 100
|||||
292 AGAGCAGGCGCCTCTGGGACAGGTCATTCCTCCGGGACAGCGCGAGCTG 341
101 AlaAspGlnGluTyrArgGlySerPheThrCysArgIleTrpGlnPheG 117
|||||
342 GCCGACCGAGGAGTACCGGGGCTCTTCCATCTGCGCATTTGGCAGTTGG 391
117 yArgTrpValGluValThrThrAspAspArgLeuProCysLeuAlaGlyA 134
|||||
392 AGCGTGGGTGGAGGTGACACAGATGACCGCTGCGCTGCGCTGCGAGGA 441
134 rGLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
|||||
442 GACTCTGTTTCTCCCGCTGCCAGAGGAGGATGTGTTCTGGCTCCCTTA 491
151 LeuGluLysValTyrAlaLysValHisGlySerTyrGluHisLeuTrpAl 167
|||||
492 CTGGAAGAGGTCTACGCCAAGGTCCATGGTCTCTACGAGCACCTGTGGGC 541
167 acGlyGlnValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGluA 184
|||||
542 CGGGCAGGTGGCGGATGCGCTTGGTGGACCTGTACCGCGCGCGCTGGCAAAA 591

```

184  rgtprasnleuylsglyvalalaaglyserglyglnglnaspargpro 200
    |||
592  gatgaacctgaaggcgctgacgaagacggcgacacagacagccca 641
    |||
201  glyatgtprgluhlsargthrctysargcglneuleuhsleuylvasagl 217
    |||
642  gggcgctggagacacagacacttctgacgactctccactgaaggacca 691
    |||
217  ncysleuillesercyscysvalleuserproargalalaarglul 234
    |||
692  gtgtctgatacgtctgctgctgacgcccagacagctgccccggagac 741
    |||
234  euglygluphehisalaphelilevalseraspheulargluleuglly 250
    |||
742  tggggagagttccatgcttctgacgtccggagctccagaggt 791
    |||
251  glnalaglylncysileleuleuleuargllieglnasnpotprgllyar 267
    |||
792  cagcgcgccagctgcatctctgctgctgacgacgaacccctggggcgc 841
    |||
267  gtagcysrtprglnglyleutrpargglnglyglnglytrpserglv 284
    |||
842  gccgtgtgtggcagggcgcttgagagaggggggtgaagggtgagccagc 891
    |||
284  alaspalaialavalasergluleuleuserglnleuglnglyglu 300
    |||
892  tagatgacagggagtagacatctgactctgctccagctccagaaaggag 941
    |||
301  pheptrpvalgluglglupheleuargglupheaspgluleuthrva 317
    |||
942  ttctgggtggagagagagagattctctcagggagtttgaacgactcacct 991
    |||
317  lgltytrprovalthrctualaglyhsleuglnserleuyltrhrglua 334
    |||
992  tggcttaccgggtcagcagagggcgccacacgcctgacacagaga 1041
    |||
334  rgleuleucyshlsthargalaleuoproglialatrpvallysglgn 350
    |||
1042  ggtgctcttccatnaccggcgctgctgctggcgctggtcgaaggcgac 1091
    |||
351  seralagllylcyсарganasnserglypheproseranprolysph 367
    |||
1092  tcagcagagagcgtgccgagacaacagcgcttccacgacaaccccaatt 1141
    |||
367  etrpheulargvalsergluprosergluvaltyrtllealavalleunla 384
    |||
1142  ctggcctggcggtctcagaaacggagtgatgattgcccgtcctgcaga 1191
    |||
384  rgserrgleuhsialaalsptpalaglyargalalaargalaleuval 400
    |||
1192  gatccagcgtgctgacggcgagctggcgacggcgcgacactggtg 1241
    |||
401  glyaspserhslsthsertrpserproalasertleprogllylshlshy 417
    |||
1242  ggtgacagatnacttctgtgagacccagacgacatcccgggcagaaccta 1291
    |||
417  rglinalavalglyleuhsleutrpulysvalglulysargargvalasn 434
    |||
1292  ccagcgctggggtctgacacttggaaagtagaagacggcggttcacatc 1341
    |||
434  euproargvalleusermetproprovalalaglythrallacyshlsala 450
    |||
1342  tgccttaggggtctgctcattgcccccggtgctggcaccgctgcacatgca 1391
    |||
451  tyraspargluvalhslsleuargcysgluleuserprogllytyrtyrle 467
    |||
1392  ttagacacggagaggtccactgctgttgacttctacccgggctactacct 1441
    |||
467  ualavalproserthrpheluleylsaspalaprogllyglupheleu 484
    |||
1442  ggctgtccccacacacttctcgaagagacggcagggaggttctctctgc 1491
    |||
484  rgvalpheserthrglargvalserleuseralalleargalavalala 500
    |||

1492  gacttcttaccggcgagcttcccttag..... 1522
    |||
501  lysasnthrthrproglyalalaleuproalaaglyglutrpgllythrva 517
    |||
1522  ..... 1522
517  lglneulargllysertrpargvalglnglnthralaglylserarg 534
    |||
1522  ..... 1522
534  snphealasertyrprothransprocysphepropheservalproglu 550
    |||
1522  ..... 1522
551  glyprogllyproargcysvalargllethrleuhslnhscysargpr 567
    |||
1523  ggccttgccccccctgctgctccatcactctgcacacagcactggcgcc 1572
    |||
567  oserasptthrgluphehisprollieglyphehisillephegluvalpro 584
    |||
1573  cagtgacacgagttccaccccatgctgctccatattctccagagttccacg 1622
    |||
584  lglnglyltyrserglnaspalaprobleuleuleuglngluproleu 600
    |||
1623  agsgtggaagagacagacagacacccacactgctgctcagagacccgctg 1672
    |||
601  leusercysvalprohlsargtyrvalaglnlvalserargleucysle 617
    |||
1673  ctgagctgctgtgcccacatctgacgcccagagagtgagccgctgctgct 1722
    |||
617  uleuproalagllythrlylsvallvalproserthrtyrleuprokspt 634
    |||
1723  cctgcttgcagagcactacaaagttgtgcttccacttacctgcccagaca 1772
    |||
634  hrtglnglyalapherthvalthrillealathrarglleasparproser 650
    |||
1773  cagagggggcgcttcacagtgacatccgaacacagatgagacagccatcc 1822
    |||
651  llehslerglnglmetleuglgnpheleuglnglvalservalme 667
    |||
1823  atttcacgcccagagagatgctggcgacgttctccaaagaggttctccgcat 1872
    |||
667  talavalmetlysthr 672
    |||
1873  ggcagtgatgaaaaacc 1888
    |||

seq name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAA27480
seq_documentation_block:
ID AAA27480 standard; cDNA; 2516 BP.
XX
XX
AC AAA27480;
XX
XX
DT 15-AUG-2000 (first entry)
XX
XX
DE Human calpain 10e cDNA.
XX
XX
KW NIDDM1: non-insulin-dependent diabetes mellitus; CAPN10 gene;
KW calpain 10; calpain 10e; diapain-1; cysteine protease;
KW chromosome 2; human; diagnosis; therapy; ss.
OS Homo sapiens.
XX
XX
FH Key Location/Qualifiers
FT CDS 42..1376
FT /tag= a
XX
XX
XX PN WO200023603-A2.
XX
XX
XX PD 27-APR-2000.
XX
XX PF 21-OCT-1999; 99WO-US24890.

```



```

384 rGserArgLeuHisAlaAlaAspTrrAlaGlyArgAlaAlaLeuVal 400
|||||
1192 GATCCAGGCGGCTGACGCGGACTGGCGGCGGCGGCGGCGGCGGCGGCTG 1241
401 GLyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHisTyr 417
|||||
1242 GGTGACAGCATACCTTCGTGGAGCCGACGACGATCCCGGCAAGACTA 1291
417 rGlnAlaValGlyLeuHisLeuTrpLys..... 426
|||||
1292 CCAGGCTGTGGTGTGCACCTCTGAAAGGGGTGTGACACTTGGACACAC 1341
426 ..... 426
1342 TGTTCCTCTCTCCCTTCATGATGTGGCCACATGATGTTCTTCTCTT 1391
426 ..... 426
1392 GCAAAAGAGTTCGTGGAGGCCCTACTGTCCAGACGCCCGCAGTTGCT 1441
426 ..... 426
1442 GGGCCACGGTGGCTTTGTGGCCACAGCTACAGAGAGACTTGCAGGCTG 1491
427 .....Val 427
1492 TGTCTGGGACAGATAGTGGCGCCAGGCGCCAGTGAAGCCCGGATGTGA 1541
428 GLyLysArgValAlaSerProArgValLeuSerMetProProValAl 444
|||||
1542 GAGAAAGCGGCGGTCAATCTGCTAGGGTCTGTCCATGCCCCCGTGGC 1591
444 aGlyThrAlaCysHisAlaTyrAspArgGlyValHisLeuArgCysGlu 461
|||||
1592 TGGCACCCGCTGCATGATACGACGCGGAGGTCCACCTGGCTTGAGAC 1641
461 euserProGlyTyrTyrLeuAlaValProSerThrPheLeuLysAspAla 477
|||||
1642 TCTCACCGGGCTACTACCTGGCTGTCCGACACCTTCTGAAGACGCG 1691
478 ProGlyGluPheLeuArgValPheSerThrGlyArgValSerLeuSe 494
|||||
1692 CCAGGGAGTCTCTCCGAGCTTCTTACGGGCGAGTCTCCCTTAG 1740
494 rAlaIleArgAlaValAlaLysAsnThrThrProGlyAlaAlaLeuPro 511
1741 ..... 1741
511 IaGlyGluTrpGlyThrValGlnLeuArgGlySerTrpArgValGln 527
1741 ..... 1741
528 ThrAlaGlySerArgAsnPheAlaSerTyrProThrAsnProCysPh 544
1741 ..... 1741
544 eProPheSerValProGlyGlyProGlyProArgCysValArgIleThr 561
|||||
1742 .....GGCCCTGGGCGCCCGCTCGCTCCGATCATCTC 1772
561 eunHisGlnHisCysArgProSerAspThrGluPheHisProIleGlyPhe 577
|||||
1773 TGCATTCAGCATGCCGCCAGTACACCGAGTTCACCCATCGGCTTC 1822
578 HisIlePheGlnValProGluGlyGlyArgSerGlnAspAlaProPhe 594
|||||
1823 CATATCTTCCAGGTCCAGAGGGTGAAGGAGCAGGACGCCGCCACT 1872
594 uLeuLeuGlnGluProLeuLeuSerCysValProHisArgTyrAlaGln 611
|||||
1873 GCTGCTCCAGAGCGGCTGCTGAGACTGCTGCCACATCGTACGCCCGC 1922
611 LuValSerArgLeuCysLeuLeuProAlaGlyThrTyrLysValValPro 627

```

```

|||||
1923 AGGTGAGCCGGCTCTGCTCTGTCGAGGACCTACAGAGTTGGCC 1972
628 SerThrTyrLeuProAspThrGluGlyAlaPheThrValThrIleAla 644
|||||
1973 TCCACTTACTCTCCGACACAGAGGGGCGCTTACAGTACCATTCGCAAC 2022
644 rArgIleAspArgProSerIleHisSerGlnGluMetLeuGlyGlnPhe 661
|||||
2023 CAGGATTCAGGCGCATTCATTCACGCCAGGAGATGTGGCGCACTTC 2072
661 euGlnGluValSerValMetAlaValMetLysThr 672
|||||
2073 TCCAGAGGCTCTCCGATGAGGAGTGTGAAACC 2107

seq_name: /SIDSL/gcdata/geneseq/geneseqn-emb1/NA2000.DAT.AAA27479
seq_documentation_block:
ID AAA27479 standard; cDNA: 2204 BP.
XX
AC AAA27479;
XX
DT 15-AUG-2000 (first entry)
XX
DE Human calpain 10d cDNA.
XX
KW NIDDM1: non-Insulin-dependent diabetes mellitus; CAPN10 gene;
KW calpain 10; calpain 10d; diaphanin-1; cysteine protease;
KW chromosome 2; human; diagnosis; therapy; ss.
XX
OS Homo sapiens.
XX
FH Key 42..1583
FT CDS /*tag= a
FT
PN W0200023603-A2.
XX
PD 27-APR-2000.
XX
PE 21-OCT-1999; 99WO-US24890.
XX
PR 21-OCT-1998; 98US-0105052.
PR 13-MAY-1999; 99US-0134175.
XX
PA (ARCH-) ARCH DEV CORP.
XX
PI Polonsky KS, Horikawa Y, Oda N, Sreenan S, Zhou Y, Otani K;
PI Hanis CL, Bell GI.
XX
DR WPI: 2000-339702/29.
XX
P-PSDB: AAY79570.
XX
PT Method for screening for type 2 diabetes mellitus comprises detecting a
PT polymorphism in a calpain encoding nucleic acid segment or a
PT protease-encoding nucleic acid segment -
PS
XX Claim 65; Page 227; 257pp; English.
XX

```

The present sequence is that of cDNA corresponding to a minor transcript of the human calpain 10 gene, CAPN10 (see AA227475). The sequence contains an open reading frame that encodes a protein of 513 amino acids (see AAY79570), designated calpain 10d. The CAPN10 gene consists of 15 exons spanning 32 kb. A complex pattern of alternative splicing generates proteins of 672, 544, 517, 513, 444, 274, 139 and 138 amino acids (see AAY79567-74), designated calpain 10a to 10h, respectively. Calpain 10d is encoded by exons 1-7, 9 and 11-13 of the gene. The invention concerns the identification of genes responsible for type 2 diabetes for use in diagnostic and therapeutic applications. A G-to-A polymorphism in Intron 3 of the CAPN10 gene in the NIDDM1 region of chromosome 2 (UCSNP-43) shows evidence for linkage to type 2 diabetes. Claimed methods for screening for a propensity for type 2 diabetes are

CC based on detection of a polymorphism in a calpain encoding nucleic
 CC acid, especially UCSNP-10 of the CAPN10 gene. Methods are also
 CC claimed for: producing calpain 10 polypeptides using calpain
 CC 10-encoding polynucleotides, including the present sequence;
 CC identifying modulators of calpain activity using calpain 10
 CC polypeptides and nucleic acids; and using these modulators to
 CC treat diabetes, in particular through the regulation of an insulin
 CC secretory response or insulin mediated glucose transport.
 XX
 SQ Sequence 2204 BP; 390 A; 672 C; 752 G; 390 T; 0 other;

alignment_scores:

Quality: 3090.50 Length: 672
 Ratio: 5.292 Gaps: 1
 Percent Similarity: 86.905 Percent Identity: 86.905

alignment_block:

US-09-768-877-2 x AAA27479 ..

Align seg 1/1 to: AAA27479 from: 1 to: 2204

1 MetArgAlaGlyArgGlyAlaThrProAlaArgLeuLeuPheArgAspAl 17
 42 ATGCGGGCGGGCGGGCGGCGAGCCCGGGAGGAGCTGTTCGGGACGC 91
 17 aAlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrProL 34
 92 CGCTTCCCGCGCGGACTCCTCGCTCTCTCGGACTGTCTACCGCGC 141
 34 euAlaGlnPheArgGluAspIleThrTrpArgArgProGlnGluIleCys 50
 142 TGGCCAGTGTCCCGGAGGACATCACCTGGAGGCGGCCCGGAGATTGT 191
 51 AlaThrProArgLeuPheProAspAspProArgGluGlyGlnValLysG 67
 192 GCCACACCCCGGCTGTTCAGATGACCCAGGAGGCGAGGTGAAGCA 241
 67 nGlyLeuLeuGlyAspCysTrpPheLeuCysAlaCysAlaAlaLeuGlnL 84
 242 GGGGCTGCTGGGGATTGCTGGTTCCTGTGCTGCGCGCGGCTGCAGA 291
 84 ySerArgHisLeuLeuAspGlnValIleProProGlyGlnProSerTrp 100
 292 AGAGCAGGACCTCTCCGACACAGTATTCCTCCGCGGAGCGAGCTGG 341
 101 AlaAspGlnGluTrpArgGlySerPheThrCysArgIleTrpGlnPheG 117
 342 GCCGACCGAGGTACCGGGGCTCCTTCACCTGTGCGATTGGCAGTTGG 391
 117 yArgTrpValGluValThrTraspAspArgLeuProCysLeuAlaGlyA 134
 392 ACGCTGGTGGAGGTGACCACAGATGACCGCTGCGCTGCTTGCAGGCA 441
 134 rGlyCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
 442 GACTCTGTTCTCCCGCTCCGAGGAGGAGATGTGTCTGGCTCCCTTA 491
 151 LeuGluLysValTrpAlaLysValHisGlySerTrpGluHisLeuTrpAl 167
 492 CTGGAAGAAGGTCTACCCCAAGTCCATGGGTCTTACGAGCCTGTGGGC 541
 167 aGlyGlnValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGluA 184
 542 CGGGCAGGTGGCGGATGCCCTGGTGACCTGACCGCGGCTGGCAGAAA 591
 184 rGTrpAsnLeuLysGlyValAlaGlySerGlyGlnGlnAspArgPro 200
 592 GATGGAACCTGAAGGCGGTAGCAGGAAGCGGAGCCAGCAGGACAGGCA 641
 201 GlyArgTrpGluHisArgThrCysArgGlnLeuLeuHisLeuLysAspG 217
 642 GCGCGCTGGGAGCACAGGACTGTTCGGCAGCTGTCTCCACCTGAAGGACCA 691

217 nCysLeuIleSerCysCysValLeuSerProArgAlaGlyAlaArgGluL 234
 692 GTCTCTGATCAGCTGCTGCTGCTCAGCCACAGCAGAGGTGCCCGGAGC 741
 234 euGlyGluPheHisAlaPheIleValSerAspLeuArgGluLeuGlnGly 250
 742 TGGGGAGTTCATGCTTCATGCTTCGGAGCTTGGGGAGCTCCAGGGT 791
 251 GlnAlaGlyGlnCysIleLeuLeuLeuArgIleGlnAsnProTrpGlyAr 267
 792 CAGCGGGCCAGTGCATCTGCTGCTGCGGATCCAGAACCCCTGGGGCCG 841
 267 gArgCysTrpGlnGlyLeuTrpArgGluGlyGlyGluGlyTrpSerGlnV 284
 842 CGGCTGCTGCGAGGGGCTCTGGAGAGAGGGGGGTGAGGGTGGAGCCAG 891
 284 aAlaAlaAlaValAlaSerGluLeuLeuSerGlnLeuGlnGluGlyGlu 300
 892 TAGATGACGGGTAGCATCTGAGCTCTGTCCAGCTCCAGGAGGGGAG 941
 301 PheTrpValGluGluGluPheLeuArgGluPheAspGluLeuThrVa 317
 942 TTTCTGGGTGGAGGAGGAGGAGTTCTCAGGGAGTTTGACGAGCTCACCGT 991
 317 lGlyTyrProValThrGluAlaGlyHisLeuGlnSerLeuTyrThrGluA 334
 992 TGCTACCCCGGTCACGAGGCGCGCCACCTGCAGAGCCTCTACACAGAGA 1041
 334 rGLeuLeuCysHisThrArgAlaLeuProGlyAlaTrpValLysGlyGln 350
 1042 GGCTGCTCTGCCATACGGGGCGCTGCTGGGGCTTGGGTCAAGGGCCAG 1091
 351 SerAlaGlyCysArgAsnAsnSerGlyPheProSerAsnProLysPh 367
 1092 TCAGCAGGAGGTGCGCGGAACACAGCGGCTTCCAGCAACCCCAATT 1141
 367 eTrpLeuArgValSerGluProSerGluValTyrIleAlaValLeuGlnA 384
 1142 CTGGCTTCCGGGCTCAGAACCCGAGTGAGGTGTACATTGCCGCTCTGCAGA 1191
 384 rGSerArgLeuHisAlaAlaAspTrpAlaGlyArgAlaArgAlaLeuVal 400
 1192 GATCCAGGCTGCACGCGCGGACTGGGCGCGCGCGCGGCGGCTGGTG 1241
 401 GlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHisTy 417
 1242 GGTGACGTATCTGCTGAGAGCCAGCGAGCATCCCGGGCAAGCACTA 1291
 417 rGlnAlaValGlyLeuHisLeuTrpLysValGluLysArgArgValAsnL 434
 1292 CCAGGCTGTGGGTCTGCACCTCTGGAAGGTAGAGAAAGCGGGGTCAATC 1341
 434 euProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAla 450
 1342 TGCTTAGGGTCTGTCTCATGCCCCCGTGGCTGGCACCGCTGCCATGCA 1391
 451 TyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTyrTyrLe 467
 1392 TACGACCGGGAGGTCCACCTGCTGTGAGCTCTCACCAGGCTACTACCT 1441
 467 uAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeuLeuA 484
 1442 GGCTGTCCCCAGCCTTCTCTGAAGAGCGCGCAGGGAGTTCTTCTGCTCC 1491
 484 rGValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAla 500
 1492 GAGTCTTCTTACCGGGCGAGTCTCCCTTAG..... 1522
 501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyGluTrpGlyThrVa 517
 1522 1522

XX Hanis CL, Bell GI;
DR WPI: 2000-339702/29.
XX P-PSDB: AAY79575, AAY79574, AAY79576.

PT Method for screening for type 2 diabetes mellitus comprises detecting a
PT polymorphism in a calpain encoding nucleic acid segment or a
PT protease-encoding nucleic acid segment -

PS Claim 65; Page 236-237; 257pp; English.

CC The present sequence corresponds to the major transcript of the
CC mouse CAPN10 gene encoding calpain 10 (see AAY79575), a novel
CC cysteine protease. Calpain is ubiquitously expressed in mouse and
CC humans. The invention concerns the identification of genes
CC responsible for type 2 diabetes for use in diagnostic and
CC therapeutic applications. A polymorphism in intron 3 of the human
CC CAPN10 gene (see AAA27475) shows evidence for linkage to type 2
CC diabetes. Claimed methods for screening for a propensity for type
CC 2 diabetes are based on detection of a polymorphism in a calpain
CC encoding nucleic acid. Methods are also claimed for: producing
CC calpain 10 polypeptides using calpain 10-encoding polynucleotides,
CC including the present sequence; identifying modulators of calpain
CC activity using calpain 10 polypeptides and nucleic acids; and using
CC these modulators to treat diabetes, in particular through the
CC regulation of an insulin secretory response or insulin mediated
CC glucose transport.

SX Sequence 2511 BP; 523 A; 708 C; 749 G; 531 T; 0 other;

Alignment_scores:
Quality: 2963.00 Length: 671
Ratio: 4.756 Gaps: 2
Percent Similarity: 92.846 Percent Identity: 81.073

Alignment_block:
US-09-768-877-2 x AAA27484 ..

Align seg 1/1 to: AAA27484 from: 1 to: 2511

1 MetrghlacllyarqglialathlrProalaahrgluleauphearqspal 17
||||| |||||
109 ATGGGCGGCGGTTCGGGGCCAGAGGCCGGCGCGGAGCTTTCGGGAGC 158
17 aaIaphPrOAlaAlAsPserSeriLeupheCyaspenseEthrPol 34
|||||
159 GGACTTCCCGCTCGGACTCCTCGCTTTACAACTGTGCAGGCTC 208
34 eulaagInPheargIuaSpIlerrThrPatargPrgInglnuIEcys 50
209 TGGCCCGATTTCGGGAGCATCACTATGGAGACGCCCAAGAATCTGT 258
51 AlAthPrOArgrleuphePrOaspsPrOatrgIugIygnValLygI 67
|||||
259 GCCAACACTCAGCTGTTCACAATTAACCATGGGAGGACAGGTGAAGA 308
67 nglYleuLEuGLyASPCysTrPheLeuCysAlAcysAlaIaleuGINt 84
309 AGGGCTGTGGGAGATTGCTGTCTCTGTGTGTCCTCGTGCCCTTAGA 358
84 yseSerarghISLeuLeuaspgInValIIePrOfroGLygnIProSeTrTP 100
|||||
359 AGAGTCAACACTCTCGGACACAGGTCTTCCTCCAGGACACCCAGGCTGG 408
101 AlaSPoGlngluTrarGrLyserPhehThrysaTglierPgInPheGI 117
:::|||||:::|||||
409 TCtGACCAcCAAtACCAAGGCTTTCACCTGTGCGATTGGCAgtTtgG 458
117 yarGTrPvalIgIuValThrThraSPaRPaRyleuProcysLeuaIGLYA 134
:::|||||
459 ACActGGAGAGAAgtACCATgATgATGATGCTGCTTGTCTTGGCCGGGA 508

134 rgLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
|||||
509 GACTCTGCTTTCCCGTGCCAGAGAGAGATGTCTGGCTTCCCTTA 558
151 LeuGluLysValTrpAlaLysValHisGlySerTrpGluHisLeuTrpAl 167
|||||
559 CTGGAAGCCCTATGCTAGGTCCTATGATGATGATGATGATGATGATG 608
167 agGlnValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGluA 184
|||||
609 AGGCCAAGTGGCAGATGCTTGTAGTGGATCTCACTGGAGCCCTGGCAGAA 658
184 rgTrpAsnLeuLysGlyValAlaGlySerGlyGlyGlnGlnAspArgPro 200
|||||
659 GGTGGAGCTTGAAGGATGTAAACGAAGCCAGCGGCCAGCAGCAGACCC 708
201 GlyArgTrpGluHisArgThrCysArgGlnLeuLeuHisLeuLysAspG 217
:::
709 AGTGGTGGGAGCACAGAACTGTGCGCAGCTACTCCACTGAAGGACCG 758
217 nCysLeuLeuSerCysValLeuSerProArgAlaGlyAlaArgGluL 234
|||||
759 GTGCTCTAATCAGCTGCTGTGCTGTAGCCCCAGAGCAGGTGCCAGGAAC 808
234 euGlyCluPheHisAlaPheIleValSerAspLeuArgGluLeuGly 250
|||||
809 TCGGAGAGTTCATGCTTATCATCTCAGATCTGCGAGGCTCAGGAGT 858
251 GlnAlaGlyCysIleLeuLeuLeuArgIleGlnAsnProTrpGlyAr 267
|||||
859 CAGACTGGCCAGGGTATCTCTGCTGCGGATTCAACCCCTGGGGCCG 908
267 gArgCysTrpGlnGlyLeuTrpArgGluGlyGlyGlyGlyTrpSerGlnV 284
|||||
909 CGGTGTGGCAGGGCTCTGAGAGAGAGAGGTGAAGGTGGAACCCAGG 958
284 alaSpAlaValAlaSerGluLeuLeuSerGlnLeuGlnGlyGlu 300
||::: |||
959 TAGAGCCAGCTAAGGAGTCTGAGCTGCTGGCCCACTCCAGGAAGGAGAG 1008
301 PheTrpValGluGluGluPheLeuArgGluPheAspGluLeuThrVa 317
|||||
1009 TTCTGGGTGAGAGAGAGAGTCTCAGGGAGTTTGATGAGGTCAACAT 1058
317 lGlyTrpProValThrGluAlaGlyHisLeuGlnSerLeuTrpThrGluA 334
|||||
1059 CGGTACCCAGTCACAGAGCGCGGCCACTACAGAGTCTCCACACAGAGA 1108
334 rgLeuLeuCysHisThrArgAlaLeuProGlyAlaTrpValLysGlyGln 350
|||||
1109 GGGTCTGTGCCATACGGCGACACTGCTGCTGCTGCTGCTGCTGCTGCTG 1158
351 SerAlaGlyGlyCysArgAsnAsnSerGlyPheProSerAsnProLysPh 367
|||||
1159 TCAGCAGAGAGGCTGCCGAACANCAAGTTGCTTTCCCTGCAACCCCAAGTT 1208
367 eTrpLeuArgValSerGluProSerGluValTrpIleAlaValLeuGlnA 384
|||||
1209 CTGGTTACGGCTCTTGGAAACCAGCAGGAGGTGTGTGTGTGTGTCTTCA 1258
384 rgSerArgLeuHisAlaAlaAspTrpAlaGlyArgAlaArgAlaLeuVal 400
|| |||
1259 GACCCCGGAGCGGCTA.....GTGGCCAGACTCGGCGACTGGCG 1299
401 GlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHisTy 417
|||
1300 GGT.....GCCAGTCTGACCGGTGACCTCCCGCAGCAAGACTA 1340
417 rGlnAlaValGlyLeuHisLeuTrpLysValLysValArgValAsnL 434
|||||
1341 CCAGGCTGTGGCCCTGCACATCTGAAGGTAGAGAAACGGAAGATCAGCC 1390
434 euProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAla 450

1391 TGCCAGAGTCTCTGTGACACCCCTGTGGCTGGCACTGCATGCCATGGC 1440
451 TyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTrpTyrIle 467
|||||
1441 TATGATCGTGAGATCCACTTGGCTTGTGAGCTCTCACAGGCTACTACCT 1490
467 uAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeuLeuA 484
|||||
1491 GGCGCTCCCTAGCACCTTTTGAAGATGTGCCAGGGCAGTCTCTGCTCA 1540
484 rgValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAla 500
|||||
1541 GAGCTCTTCTCACTGGGAAATCTCCCTCAGTGCCTCAGGCTGGCCACC 1590
501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyGluTrpGlyThrVa 517
|||||
1591 AAGGGTGCATCGCTCGGAACAGCCCTCGCTGCAGGCGAGTGGGAGACTGT 1640
517 lGlnLeuArgGlySerTrpArgValGlyGlnThrAlaGlyGlySerArgA 534
|||||
1641 GCAGTGGCAGGGCTGCTGGAGAGCTGCCAGACAGCTGGGGCGCAGCAGA 1690
534 snPheAlaSerTrpProThrAsnProCysPheProPheSerValProGlu 550
|||||
1691 ACTTTGCCCTTTACCCCTGCAATCCCTGCTCCCTTCTCTCTCTCTGAG 1740
551 GlyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArgPr 567
|||
1741 GGTGCTGGCCCGCTGATCATCGTATCACCTACACAACTGCGCGCT 1790
567 oSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValProG 584
|||||
1791 CAGTGCACCCAGCTGCACCCCATTTGGTTTCCATGTTCTTCAAGTTCCAG 1840
584 luGlyGlyArgSerGlnAspAlaProProLeuLeuLeuGlnProLeu 600
||::: |||
1841 CAGACGGTGCAGAACCCAGGAGCGTGTTCCTGCTGCTCCAGGAGCCACTG 1890
601 LeuSerCysValProHisArgTrpAlaGlnGlnValSerArgLeuCysLe 617
|||||
1891 CTAAGCTGTGTACCATCGCTACGCCAGGAAGTACGCCGCTCTGCT 1940
617 uLeuProAlaGlyThrTrpLysValValProSerThrTrpLeuProAspT 634
|||
1941 CTTTCTGTGGGGAACACTACAGATGTTCCTCCACCTACCTACCTGCGCAGATA 1990
634 hrGluGlyAlaPheThrValThrIleAlaThrArgIleAspArgProSer 650
|||||
1991 CAGAGGTACCTTACGGTTAACCATAGCAACCAAGATCGATAGGAGTCC 2040
651 IleHisSerGlnGluMetLeuGlyGlnPheLeuGlnGluValSerValMe 667
|||||
2041 ATCCACAGCCAGGAGATGCTGGCCAGCTGCTCCAGAGGTCTCTCTTAT 2090
667 tAlaValMetLys 671
|||||
2091 GGCAGTGATGAAA 2103

seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:AAS97171

seq_documentation_block:

ID_AAS97171 standard; cdna; 1542 BP.

XX AAS97171;

XX XX

DT 26-FEB-2002 (first entry)

XX Human cysteine protease partial DNA sequence #9.

XX Human; protease; PCR primer; cytostatic; immunomodulator; cardiant;
KW vasotrophic; antimigraine; analgesic; endocrine; nootropic; tranquiliser;
KW hypertensive; hypotensive; neuroleptic; neuroprotective; anabolic;

51	AlaThrProArgLeuPheProAspProArgGluGlyGlnValIysGlu	67
151	GGCACACCCCGCGCTGTTTCCAGATGACCCACGGGAAAGCGATGAAGCA	200
67	nglyLeuLeuGlyAspCysTrpPheLeuCysAlaCysAlaAlaLeuGln	84
201	GGGGCTGCTGGGGGATTTGCTGTTCTGTGGCTGGCCGCGCTGGAGA	250
84	YSSerArgHisLeuLeuAspGlnValIleProProGlyGlnProSerTrp	100
251	AGACACAGGACACTCTGTGGACAGGTCTATTCCTCGGGAGACCGAGCTGG	300
101	AlaAspGlnGluTrpArgGlySerPheThrCysArgIleLeuGlnPheGlu	117
301	GGCGACACGAGAGTACCGGGGCTCTTACCTGTGCGCATTTGGCAAGTTGG	350
117	YArgTrpValGluValIleThrThrAspAspArgLeuProCysAlaGlu	134
351	ACGCTGGGTGGAGGTGACCAAGATGACCCGCTGGCCGCTTGACAGGA	400
134	rgLeuCysPheSerArgCysGluArgGluAspValPheTrpLeuProLeu	150
401	GACTCTGTTCCTCCGCTGCGACAGGGAGGATGTTGTTGGCTCCCTTA	450
151	LeuGluIysValIleTrpAlaIysValHisGlySerTrpGlnHisLeuTrpAl	167
451	CTGGAAAGGTCTACGCGCAAGTCCAGTCCAGAGCACTGTGGGC	500
167	agGlyGlnValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGlu	184
501	CGGGCAGGTGGCGGATGCCCTGGTGAGCACTGACCGGGGCTTGCAAAA	550
184	rgTrpAsnLeuIysGlyValAlaGlySerGlyGlyGlnGlnAspArgPro	200
551	GATGGAACTCGAAAGGGGCTGACGAGAAACGGAGCCAGCAGCAAGCCA	600
201	GlyArgTrpGlnHisArgTrpCysArgGlnLeuLeuHisLeuIysAspGlu	217
601	GGCGCGTGGGAGACAGAGCACTTGTCCGCACTCTCCACTGAAAGACCA	650
217	ncYsLeuLeuSerCysCysValLeuSerProArgAlaGlyAlaArgGlu	234
651	GTCGTGTGATAGCTGTGCTGCTGCACGCCACAGACAGAGTGCCCGGAGC	700
234	euGlyGluPheHisAlaPheIleValSerAspLeuArgGluLeuGlnGly	250
701	TGGGGGAGTTCACATGCCCTTCAATTGTCTCGGACCTCGGGGAGCTCCAGGCT	750
251	GlnAlaGlyGlnIysIleLeuLeuLeuArgIleGlnAsnProTrpGluTrp	267
751	CAGCGGGGCGCAGTGCATCTCGCTGCTGCGGATCCAAACCCCTGGGGCCG	800
267	gaArgCysTrpGlnGlyLeuTrpArgGluGlyGlyGlnGlyTrpSerGlnY	284
801	GGCGTCTCTGACAGGGGCTCTGAGAGAGGGGGGTGAAGGTTGAGGCCAGG	850
284	alaAspAlaAlaValAlaSerGluLeuLeuSerGlnLeuGlnGluGlyGlu	300
851	TAAATCCAGCGGTAGAGATCTGACACTCTCTCCACGCTCCAGAGAAAGGGAG	900
301	PheTrpValGluGluGlnGluPheLeuArgGluPheAspGluLeuThrPa	317
901	TTCCTGGGTGGAGAGAGAGAGTTCTCTCAAGGAGTTTGACGAGCTCAAGCT	950
317	IleGlyTrpProValIleThrGluAlaGlyHisLeuGlnSerLeuTrpThrGluA	334
951	TGGCTACCGCGGTACGAGAGCGCGGCACCTGCAGAGCCTCTACACAGAGA	1000
334	rgLeuLeuCysHisThrArgAlaLeuProGlyAlaLeuValIysGlyGln	350
1001	GGCTGCTCTGACCATACGCGGGGCTCTCTGGGGGCTGGGTCAAGGGCCAG	1050

351 SerAlaGlyGlyCysArgAsnAsnSerGlyPheProSerAsnProLysPh 367
1051 TCACGAGAGGCTGCCGGAACAACAGCGGCTTTCCACGACACCCCAATT 1100
367 eTTPLeuArgValSerGluProSerGluValTyrIleAlaValLeuGlnA 384
1101 CTGGCTCGGGTCTCAGAACCGAGTAGGTGTACATGTCGCTCGTCGAGA 1150
384 rGSerArgLeuHisAlaAlaAspTrpAlaGlyArgAlaArgAlaLeuVal 400
1151 GATCCAGGCTGCACGGCGGCGGACTGGCAGCGCGGCGGCGGCGGCTGGT 1200
401 GlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHisTy 417
1201 GGTGACAGTCATCTTGTGTGAGCCCGCAGCGAGCATCCCGGCGCAGCACTA 1250
417 rGlnAlaValGlyLeuHisLeuTrpLysValGluLysArgArgValAsnL 434
1251 CCAGGCTGTGGGTCTGCACCTCTGGAGGAGTAGAGCGCGGGGTCAATC 1300
434 euProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAla 450
1301 TGCTAGGGTCTCTCATGCCCGCGGCTGGCTGGCACCCTGCGCATGCA 1350
451 TyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTyrTyrLe 467
1351 TACGACCGGAGGTCCACCTGCTGTGTGAGCTCTCACCGGCTACTACCT 1400
467 uAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeuLeuA 484
1401 GGTGTCTCCAGCACCTTCTGAAGAGCGCGCGGAGGTCTCTGCTCC 1450
484 rGValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAla 500
1451 GAGTCTCTCTACCGGCGGAGTCTCCCTTAG 1481
501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyGluTrpGlyThrVa 517
1481 1481
517 lGlnLeuArgGlySerTrpArgValGlyGlnThrAlaGlyGlySerArgA 534
1481 1481
534 snPheAlaSerTyrProThrAsnProCysPheProPheSerValProGlu 550
1481 1481
551 GlyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArgPr 567
1481 1481
567 oSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValProG 584
1482GTCCAG 1488
584 luGlyGlyArgSerGlnAspAlaProProLeuLeuGlnGluProLeu 600
1489 AGGGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1538
601 Leu 601
1539 CTG 1541

seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AA13883
seq_documentation_block:
ID AA13883 standard; cDNA; 2180 BP.
XX AA13883;
AC
XX
DT 29-NOV-2001 (first entry)
XX

DE XX Human calpain-like protease, 18036 cDNA.
KW Calpain-like protease 18036; spleen disorder; splenomegaly; lung;
KW adult respiratory distress syndrome; colon; liver; jaundice; brain;
KW idiopathic inflammatory bowel disease; cerebrovascular disease;
KW acute meningitis; Alzheimer's disease; T-cell; transplant rejection;
KW systemic lupus erythematosus; skin; seborrhoeic keratosis; leukaemia;
KW haematopoietic stem cell; heart; myocardial infarction; atherosclerosis;
KW kidney; acute proliferative glomerulonephritis; urolithiasis; apoptosis;
KW ischaemia; neurodegenerative disease; demyelinating disease; injury; EAE;
KW experimental allergic encephalomyelitis; multiple sclerosis; spinal cord;
KW LGMD2A muscular dystrophy; proliferative disorder; cancer; gene therapy;
human; chromosome 2; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 183..1736
FT /*tag= a
FT /product= "Calpain-like protease, 18036"
FT /note= "This region is claimed as SEQ ID NO: 3 in
FT claim 1"
XX
PN WO200164919-A2.
XX
XX 07-SEP-2001.
XX
XX 26-FEB-2001; 2001WO-US06110.
XX
XX 28-FEB-2000; 2000US-0185333.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Kapeller-Libermann R;
XX
XX WPI: 2001-565507/63.
XX P-PSDB; NAE09756.
XX
XX New calpain-like protease polypeptides and polynucleotides for
XX diagnosing, treating seven-transmembrane protein/receptor-related
XX disorders and to identify modulators of therapeutic use
XX
XX Claim 1; Fig 1; 114pp; English.
XX
XX The invention relates to polypeptide and polynucleotide of calpain-like
XX protease, 18036. The polypeptide and polynucleotide of the invention are
XX useful for treatment and diagnosis of calpain-like protease-mediated
XX disorders. These disorders include disorders of spleen (e.g.
XX splenomegaly), lung (e.g. adult respiratory distress syndrome), colon
XX (e.g. idiopathic inflammatory bowel disease), liver (e.g. jaundice),
XX brain (e.g. traumatic brain injury, cerebral oedema, cerebrovascular
XX disease, acute meningitis, Alzheimer's disease), T-cells (e.g. transplant
XX rejection, systemic lupus erythematosus), skin (e.g. seborrhoeic
XX keratosis), haematopoietic stem cells (e.g. leukaemias), heart (e.g.
XX myocardial infarction, atherosclerosis), kidney (e.g. acute proliferative
XX glomerulonephritis, urolithiasis), thymus (e.g. acute proliferative
XX epidiolymis, endometrium, prostate, thyroid, skeletal muscle, pancreas,
XX small intestine, disorders related to reduced platelet number, bone,
XX ovary, pain and infectious disorders. The molecules of the invention are
XX also useful for diagnosing and treating disorders associated with
XX perturbed cellular growth and differentiation, exercise-induced injury
XX and repair, apoptosis, ischaemia, neurodegenerative diseases,
XX demyelinating diseases including experimental allergic encephalomyelitis
XX (EAE) and multiple sclerosis. LGMD2A muscular dystrophy, spinal cord
XX injury, proliferative and differentiative disorders e.g. cancer and
XX renal cell death associated with diverse toxicants. The present sequence
XX is human calpain-like protease, 18036 cDNA. The 18036 gene is located on
XX chromosome 2 between D2S140 and D2S2338.
XX
XX Sequence 2180 BP; 395 A; 651 C; 764 G; 370 T; 0 other;
XX
XX alignment_scores:

```
alignment_block:
US-09-768-877-2 x AAD13883
```

1	MeATrAlaGlyIyrGgiIyAlaTrnProAlaArgIuIueuPhaArgAspAl	17
183	ATCGGGGGGGGGGGGGGGAGAGCCCGGACAGGAGAGCTGTTCGGGACGC	223
17	aAlaPhrOAlaAlaAspSerSerLeuPhcysAspLeuSerThrProL	34
233	CGGCTTCCCGCGGACCTCCGCTCTTCGAGACTTGTACGGCGC	282
34	euaIaGlnPheArgGluAspIleThrTrpArgArgProGlnGluIleCys	50
283	TGGCCAGTTCGCGAGACATACGTGAGAGGGCCCCAGGAGATTGT	332
51	AlaThrProArgLeuPheProAspAspProArgGluGluValIlysgI	67
333	GCACACCCCGGCTTTCAGATGACCCAGGGAGAGAGAGGTGAAGCA	382
67	nAlaLeuLeuGluGlyAspCysTrpPheLeuGlyAlaCysAlaAlaLeuGlnL	84
383	GGGGCTGCGGGGGGATTGCTGTGCTCTGTGCTCGCGCGCTCAGA	432
84	ySerArgHisLeuLeuAspGlnValIleProProGlyGlnProSerTrp	100
433	AGACGACGACCTCCGAGACAGGTATTCCTCCGGGAGACCGACCTGG	482
101	AlaAspGlnGluTrpArgIlySerPheThrCysArgIleGtrGlnPheG	117
483	GGCGACAGGAGTACCGGGGGCTTCACCTGCGCATTTGGCAGTTTGG	532
117	yArgTrpValGluValThrThAspAspArgLeuProCysLeuAlaGluA	134
533	ACGCTGGGGGAGGTGACACAGATGACCGCTCGCTGCTCGAGGGA	582
134	rGLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu	150
583	GACTGTGTTCTCCCGCTCCAGAGGAGAGATGTCTTCGGCTCCCTTA	632
151	LeuGluLysValTrpAlaLysValHisGlySerTrpGlnHisLeuTrpAl	167
633	CTGGAAAAGGTACGCCAAGTCCATGGGTCTTACAGACACCTGTGGGC	682
167	aGlyGlnValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGluA	184
683	CGGGCAGGTGGGGAGTCCCTGTGGACTGACCGGGCGCTGGGAGAA	732
184	rGTrpAsnLeuLysIlyValAlaGlySerGlyGlyGlnAspArgPro	200
733	GATGCAACCTTAAGGGCGTACGAGAAACGGAGCCAGCAGAGACAGCA	782
201	GlyAlaGTrpGluHisArgThrCysArgGlnLeuLeuHisLeuLysAspG	217
783	GGCCCGCTGGAGACACAGGACTTGTGGCAGCTGCTCCACTGAAGACCA	832
217	nCysLeuLieserCysCysValLeuSerProArgAlaGluAlaArgGluL	234
833	GTrpCTGATGACGTGCTGCTGCTTACGCCCAAGACAGAGTCCCGGAGAC	882
234	eugIyGluPheHisAlaPheIleValSerAspLeuArgGluIueuGlnGly	250
883	TGGGGAGATTCCATGGCTTCAATTGCTCGAGACTCGGAGACTCCAGGGT	932
251	GlnAlaGlyGlnCysIleLeuLeuLeuArgIleGlnAsnProTrpGlyArg	267
933	CAGGGGGGACGATCCTCTCTGTGTGGGATTCAGAAACCTCGGGGCGG	982

```
1461 .....|
584 luGlyGlyArgSerGlnAspAlaProProLeuLeuLeuGlnGluProLeu 600
1468 AGGGTGAAGAGCCAGGACGACCCCACTGCTGTCAGAGCGCTG 1517
601 LeuSerCysValProHisArgTyrAlaGlnGluValSerArgLeuCysLe 617
1518 CTCAGCTGCTGCCACATCGCTACGCCAGGAGGTGAGCGGCTCTGCT 1567
617 uLeuProAlaGlyThrTyrLysValValProSerThrTyrLeuProAspT 634
1568 CTGCTGCTGGGGCACCCTACAGGTTGTGCCCTCCACCTACCTGCGGGA 1617
634 hrGlulAlaPheThrValThrIleAlaThrArgIleAspArgProSer 650
1618 CAGAGGGGCTTCACGTGACCATCGNACAGGATGACAGGCAATCC 1667
651 IleHisSerGlnGluMetLeuGlyGlnPheLeuGlnGluValSerValMe 667
1668 ATTCACAGCCAGGAGATGCTGGCCAGTTCTTCCCAAGAGGTCTCGTCAT 1717
667 tAlaValMetLysThr 672
1718 GCCAGTGATGAATCC 1733
```

seq_name: /SIDS1/qcgdata/geneseq/geneseq-emb1/NA2000.DAT:AAA27478

seq_documentation_block:
ID AAA27478 standard; CDNA; 2001 BP.

XX AC AAA27478;
XX DT 15-AUG-2000 (first entry)
XX DE Human calpain 10c cDNA.
XX KW NIDDM1; non-insulin-dependent diabetes mellitus; CAPN10 gene;
XX KW calpain 10; calpain 10c; diapaen-1; cysteine protease;
XX KW chromosome 2; human; diagnosis; therapy; ss.
XX OS Homo sapiens.

Key	Location/Qualifiers
CDS	42..1595
FT	/*tag- a
PN	WO200023603-A2.
PD	27-APR-2000.
XX	21-OCT-1999; 99WO-US24890.
XX	21-OCT-1998; 98US-0105052.
PR	13-MAY-1999; 99US-0134175.
XX	(ARCH-) ARCH DEV CORP.

PI Polonsky KS, Horikawa Y, Oda N, Sreenan S, Zhou Y, Otani K;
PI Hanis CL, Bell GI;
XX WPI: 2000-339702/29.
DR P-PSDB; AAY79569.

XX Method for screening for type 2 diabetes mellitus comprises detecting a
PT polymorphism in a calpain encoding nucleic acid segment or a
PT protease-encoding nucleic acid segment -

XX Claim 65; Page 224-225; 257pp; English.
XX The present sequence is that of cDNA corresponding to a
CC transcript of the human calpain 10 gene, CAPN10 (see AA227475).

CC The sequence contains an open reading frame that encodes a protein
CC of 517 amino acids (see AAY79569), designated calpain 10c. The
CC CAPN10 gene consists of 15 exons spanning 32 kb. A complex pattern
CC of alternative splicing generates proteins of 672, 544, 517, 513,
CC 444, 274, 139 and 138 amino acids (see AAY79567-74), designated
CC calpain 10a to 10h, respectively. Calpain 10c is encoded by exons
CC 1-7 and 11-13 of the gene. It is readily detectable in many
CC tissues, including skeletal muscle and islets. The invention
CC concerns the identification of genes responsible for type 2 diabetes
CC for use in diagnostic and therapeutic applications. A G-to-A
CC polymorphism in intron 3 of the CAPN10 gene in the NIDDM1 region of
CC chromosome 2 (UCSNP-43) shows evidence for linkage to type 2
CC diabetes. Claimed methods for screening for a propensity for type
CC 2 diabetes are based on detection of a polymorphism in a calpain
CC encoding nucleic acid, especially UCSNP-10 of the CAPN10 gene.
CC Methods are also claimed for: producing calpain 10 polypeptides
CC using calpain 10-encoding polynucleotides, including the present
CC sequence; identifying modulators of calpain activity using calpain
CC 10 polypeptides and nucleic acids; and using these modulators to
CC treat diabetes, in particular through the regulation of an insulin
CC secretory response or insulin mediated glucose transport.
XX

SQ Sequence 2001 BP; 361 A; 602 C; 690 G; 348 T; 0 other;

alignment_scores:

Quality: 2715.50 Length: 672
Ratio: 5.252 Gaps: 1
Percent Similarity: 76.935 Percent Identity: 76.786

alignment_block:
US-09-768-877-2 x AAA27478

Align seg 1/1 to: AAA27478 from: 1 to: 2001

1 MetArgAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
42 ATGCGGGGGCGCGGGCGGCGACCGCGAGGAGCTGTTCCGGGAGCG 91
17 aAlaPheProAlaAlaSerSerLeuPheCysAspLeuSerThrProL 34
92 CGCTTCCCCCGCGGACTCTCTGCTTCTGCGACTTGTCTACGCGCG 141
34 euAlaGlnPheArgGluAspIleThrTrpArgArgProGlnGluIleCys 50
142 TGGCCCGAGTTCGCGAGGACATCATCGTGGAGCGCGCCCGAGAGATTGT 191
51 AlaThrProArgLeuPheProAspAspProArgGluGlyGlnValLysG 67
192 GCCACACCCCGGCTGTTCCAGATGACCCAGGGAAGGCGAGGTGAAGCA 241
67 nGlyLeuLeuGlyAspCysTrpPheLeuCysAlaCysAlaAlaLeuGlnL 84
242 GGGGCTGCTGGGGGATTGCTGGTTCTGTGTGCTGCGCGCGCTGCAGA 291
84 ysSerArgHisLeuLeuAspGlnValIleProProGlyGlnProSerTrp 100
292 AGACAGCAGCACCCTCTGGACCGGTCATTCCTCGGACAGCCAGCTGG 341
101 AlaAspGlnGluTyrArgGlySerPheThrCysArgIleTrpGlnPheG 117
342 GCCGACAGGAGTACCGGGGCTCTTCCACCTGTCCGATTTGGCAGTTGG 391
117 YArgTrpValGluValThrThrAspAspArgLeuProCysLeuAlaGlyA 134
392 ACGCTGGGTGGAGTGACACAGATGACCGCTGCGCTTGCAGGGA 441
134 rgLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
442 GACTCTGTTTCTCCCGCTGCCAGAGGAGGATGTGTTCTGCTCCCTTA 491
151 LeuGluLysValTyrAlaLysValHisGlySerTyrGluHisLeuTrpAl 167


```

492 CTGAAAAGTCTACGCCAAGTCCATGGTCTTACGAGCACTCTGGGC 541
167 aglyGlnValAlaAspAlaLeuValAspLeuThrGlyLeuAlaGln 184
|||||
542 CGGCGAGTGGCGGATGCCCTGGTGAGCTGACCGCGCGCTGGAGAAA 591
184 rgtPAspLeuLysGlyValAlaGlySerGlyGlyGlnAspArgPro 200
|||||
592 GATGAACCTGAAGGCGCTAGCAGAAAGGAGCCAGCAGACAGGCCA 641
201 gLyArgTrpGlnHisArgThrCysArgGlnLeuHisLeuLysAspG 217
|||||
642 GCGCGCTGGAGCACAGACACTTGTGGGACCTGCTCCACTGAAGACCA 691
217 nCysLeuLleSerCysCysValLeuSerProArgAlaGlyAlaArgGln 234
|||||
692 GTGTCTGATCAGTGTGCTGCTCAGCCCGAGAGAGGTCCCGGGAGC 741
234 euGlyGlnPheHisAlaPheIleValSerAspLeuArgGlnLeuGlnGly 250
|||||
742 TGGGGAGATTCCATGCCCTTCTGCGGACCTGCGGAGCTCCAGGGGT 791
251 GlnAlaGlyGlnCysIleLeuLeuLeuArgIleGlnAsnProTrpGlyArg 267
|||||
792 CAGCGCGGCCAGTGCATCTGCTGTGGATCCAGAACCCCTGGGGCCG 841
267 gArgCysTrpGlnGlyLeuTrpArgGlnGlyGlyGlyTrpSerGln 284
|||||
842 GCGGTGCTGGCAGGGGCTGGAGAGAGGGGGTGAAGGGTGGAGCCAGG 891
284 alaAspAlaValAlaSerGlyLeuLeuSerGlnLeuGlnGlyGln 300
|||||
892 TAGATGAGCGGCTGACATCTGAGCTCTGCCAGCTCCAGAGAGGGAG 941
301 PheTrpValGlnGlnGlnGlnPheLeuArgGlnPheAspGlnLeuTrpVal 317
|||||
942 TTCTGGTGGAGGAGAGAGATTCTCAGGAGATTGACGAGCTCACCGT 991
317 LgIlyTrpValTrpGlnAlaGlnHisLeuGlnSerLeuTrpThrGln 334
|||||
992 TGGCTACCGGTCACGAGGCGGCGCCAGCAGAGCTTACACAGAGA 1041
334 rGlyLeuLeuCysHisTrpArgAlaLeuProGlyAlaTrpValLysGlyGln 350
|||||
1042 GGGTGGCTGCTCCATACCGGGCGCTGGGCGCTGGGCTCAAGGGCCAG 1091
351 SerAlaGlyLysArgAsnAsnSerGlyPheProSerAsnProLysPhe 367
|||||
1092 TCGACGAGAGGCTGCGGAACAACAGCGGCTTCCAGCAACCCCAATT 1141
367 eTrpLeuArgValSerGlnProSerGlnValTyrIleAlaValLeuGln 384
|||||
1142 CTGGCTGGGGGTCTCAAAACGAGTGAAGGTGATTCCTGCTGACGA 1191
384 rGSerArgLeuHisAlaAlaAspTrpAlaGlyArgAlaArgAlaLeuVal 400
|||||
1192 GATCCAGGCTGCACGCGGCGGACTGGGCAAGCGGCGCGGCACTGGTG 1241
401 GLyAspSerHisTrpSerTrpSerProAlaSerIleProGlyLysHisTrp 417
|||||
1242 GGTGACAGTCAATCTTGTGTGAGCCAGGAGCATCCCGGCAAGCACTA 1291
417 rGlnAlaValAlaGlyLeuHisLeuTrpLysValGlnLysArgArgValAsn 434
|||||
1292 CCGGGCTGTGGGTCTGCACCTCTGGAGG..... 1319
434 euProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAla 450
|||||
1319 ..... 1319
451 TyrAspArgGlnValHisLeuArgCysGlnLeuSerProGlyTyrTyrIle 467
1319 ..... 1319

```

```

467 uAlaValProSerThrPheLeuLysAspAlaProGlyGlnPheLeuAla 484
|||||
1319 ..... 1319
484 rGValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAla 500
|||||
1319 ..... 1319
501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyGlnTrpGlyThrVal 517
|||||
1319 ..... 1319
517 LglnLeuArgGlySerTrpArgValGlyGlnThrAlaGlyGlySerArg 534
|||||
1319 ..... 1319
534 snPheAlaSerTyrProThrAsnProCysPheProPheSerValProGln 550
|||||
1319 ..... 1319
551 GlyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArgPr 567
|||||
1319 ..... 1319
567 oSerAspThrGlnPheHisProIleGlyPheHisIlePheGlnValProG 584
|||||
1320 ..... GTCCAG 1326
584 LngIlyGlyArgSerGlnAspAlaProProLeuLeuGlnGlnProLeu 600
|||||
1327 AGGTTGAGAGAGCAGACGACGACCCCACTGCTGCTGCGAGAGCCGCTG 1376
601 LeuSerCysValProHisArgTyrAlaGlnGlnValSerArgLeuCysLe 617
|||||
1377 CTGAGCTGCTGCCACATCGCTACGCCCGCAGAGGTGAGCGGCTGCTGCT 1426
617 uLeuProAlaGlyThrTyrLysValValProSerThrTyrLeuProAspTr 634
|||||
1427 CTTGCTGTCAGGCACTTCAAGGTTGTGCCCTCCACTGCTGCGGACA 1476
634 hrcGlnGlyAlaPheThrValThrIleAlaThrArgIleAspArgProSer 650
|||||
1477 CAGAGGGGGCTTCACAGTACCATGCGACACAGAGTTGACAGGCCATCC 1526
651 IleHisSerGlnGlnLeuLeuGlyGlnPheLeuGlnGlnValSerValMe 667
|||||
1527 ATTCAACAGCCAGAGATGCTGGGCGCAAGTTCTCCAGAGAGTCCCGCAT 1576
667 tAlaValMetLysThr 672
1577 GGCAGTGATGAAAAAC 1592

```

seq_name: /SIDS1/gcgdata/geneseq/geneseqn_emb1/NA2001A.DAT.AAH17182

seq_documentation_block:

ID AAH17182 standard; cDNA; 1534 BP.

AAH17182;

26-JUN-2001 (first entry)

Human cDNA sequence SEQ ID NO:16542.

Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

Homo sapiens.

EP1074617-A2.

PN

XX 07-FEB-2001.

PD

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX (HELI-) HELIX RES INST.
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WPT: 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs
 XX
 PS Claim 8; SEQ ID 16542; 2537pp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 1534 BP; 286 A; 467 C; 507 G; 274 T; 0 other;

alignment_scores:
 Quality: 2487.50 Length: 511
 Ratio: 5.248 Gaps: 1
 Percent Similarity: 92.759 Percent Identity: 92.759
 alignment_block:
 US-09-768-877-2 x AAH17182
 Align seg 1/1 to: AAH17182 from: 1 to: 1534
 144 AspValPheTrpLeuProLeuLeuGluLysValTyrAlaLysValHisG1 160
 2 GATGGTTCCTGGCTCCCTTACTGAAAGAGCTACGCCAAGGTCATGG 51
 160 ySerTyrGluHisLeuTrpAlaGlyGlnValAlaAspAlaLeuValAspL 177
 52 GTCTTACGAGCAGCTGTGGCGCGGAGGTGGCGATGCCCTGGTGGACC 101
 177 euThrGlyLeuAlaGluArgTrpAsnLeuLysGlyValAlaGlySer 193
 102 TGACCGCGGCTGGCAAGAGTGAACCTGAAGGGCGTAGCAGGAAGC 151
 194 GlyGlyGlnGlnAspArgProGlyArgTrpGluHisArgThrCysArgG1 210
 152 GGAGCCAGCAGGACAGGACAGGCGCTGGGAGGACACAGGACTTGTGGCA 201

210 nLeuLeuHisLeuLysAspGlnCysLeuIleSerCysCysValLeuSerP 227
 202 GCTGCTCCACCTGAGGACCACTGCTGATCAGCTGCTCGGTCTCAGCC 251
 227 roArgAlaGlyAlaArgGluLeuGlyGluPheHisAlaPheIleValSer 243
 252 CCAGAGCAGGTGCCGGGAGCTGGGGAGTTCATGCCTTCATTGCTCG 301
 244 AspLeuArgGluLeuGlnClyGlnAlaGlyGlnCysIleLeuLeuLeu 260
 302 GACCTGCGGGAGCTCAGGGTCAGCGGGCCAGTGCATCTCTGCTGCG 351
 260 gileGlnAsnProTrpGlyArgCysTrpGlnGlyLeuTrpArgGlu 277
 352 GATCAGAACCCCTGGGCGCGGCTGTCGAGGGCTCTGGAGAGAGG 401
 277 lyGlyGluGlyTrpSerGlnValAspAlaAlaValAlaSerGluLeu 293
 402 GGGGTGAAGGTGGAGCCAGGTAGATGCAGCGGTAGCATCTCAGCTCT 451
 294 SerGlnLeuGlnGlyGluPheTrpValGluGluGluPheLeu 310
 452 TCCAGCTCCAGGAAGGGAGTTCGGGTGGAGGAGGAGTTCCTCAG 501
 310 gGluPheAspGluLeuThrValGlyTyrProValThrGluAlaGlyHisL 327
 502 GGAGTTTGAGGAGCTCAGCTGGCTACCGGTTCAGGAGGCGGCCACC 551
 327 euGlnSerLeuTyrThrGluArgLeuLeuCysHisThrArgAlaLeuPro 343
 552 TGCAGAGCTCTACACAGAGAGGCTGCTGCATACGCGGCGCTGCT 601
 344 GlyAlaTrpValLysGlyGlnSerAlaGlyGlyCysArgAsnAsnSerG1 360
 602 GGGGCTGGTCAAGGGCAGTCAAGAGGCTGCCGGACACACAGCGG 651
 360 yPheProSerAsnProLysPheTrpLeuArgValSerGluProSerGluV 377
 652 CTTTCCAGCAACCCAAATCTGCTGGGTCTCAGACCGAGTGGAG 701
 377 alTyrIleAlaValLeuGlnArgSerArgLeuHisAlaAlaAspTrpAla 393
 702 TGTACATTCGCGTCTCGAGAGATCCAGGCTGCACGCGCGGACTGGCA 751
 394 GlyArgAlaArgAlaLeuValGlyAspSerHisThrSerTrpSerProAl 410
 752 GGCGGGCGCGGCACTGGTGGTGACAGTCACTTCGTGGAGGCCAGC 801
 410 aserIleProGlyLysHisTyrGlnAlaValGlyLeuHisLeuTrpLysV 427
 802 GAGCATCCCGGCGAAGCACTACCGAGTGTGGGTCTGCACCTCTGGAAG 851
 427 alGluLysArgArgValAsnLeuProArgValLeuSerMetProProVal 443
 852 TAGAGAGCGGGGTCAATCTGCTAGGTCTCTGCATGCCCGCCCGTG 901
 444 AlaGlyThrAlaCysHisAlaTyrAspArgGluValHisLeuArgCysG1 460
 902 GCTGGCACCGGTGCTCATATGATACAGCGGAGGTCCACCTGCGTGTGA 951
 460 uLeuSerProGlyTyrTyrLeuAlaValProSerThrPheLeuLysAspA 477
 952 GCTCTACCGGGCTTACCTGCTGGCTGTCCCGACACCTTCTCTGAGGAG 1001
 477 laProGlyGluPheLeuLeuArgValPheSerThrGlyArgValSerLeu 493
 1002 CGCCAGGGAGTTCCTGCTCGAGTCTTCTACCGGGGAGTCTCCTT 1051
 494 SerAlaIleArgAlaValAlaLysAsnThrThrProGlyAlaAlaLeuPr 510
 1052 AGCCGCCATCAGGGCAGTGGGCCAAGAACACACCCCGGGGCGAGCCTGCC 1101
 510 oAlaGlyGluTrpGlyThrValGlnLeuArgGlySerTrpArgValGlyG 527

```

|||||
1102 TCGGGGGAGTGGGGACCTGTCAGCT ..... 1128
527 InrHrAlaGlyGlySerArgAspPheAlaSerTyrProThrAspProCys 543
1128 ..... 1128
544 PheProPheSerValProGlyGlyProGlyProArgCysValArgIlePhe 560
1129 .....CGCTCGTCGCGATCACC 1145
560 rLeuHisGlnHisCysArgProSerAspThrGluPheHisProIleGlyP 577
|||||
1146 TCTGATCAGACACTGCCGCCAGTACACCGAGTTCACCCCATCGGCT 1195
577 heHsIlePheGlnValProGluGlyArgSerGlnAspAlaProPro 593
|||||
1196 TCCATATCTTCCAGTCCACAGAGGTGTGAAGAGCAGACGCCACCCCA 1245
594 LeuLeuLeuGlnGluProLeuLeuSerCysValProHisArgTyrAlaG 610
|||||
1246 CTGCTGCTGCAGAGCCGCTCTGAGCTGCCTGCACATGCTACGCCCA 1295
610 nGluValSerArgLeuCysLeuLeuProAlaGlyThrTyrIlyValValP 627
|||||
1296 GGAAGGTAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1344
627 rOSeRThrTyrLeuProAspThrGluGlyAlaPheThrValThrIleAla 643
|||||
1345 CCTCCACTCTACTGCTGCCGACACAGAGGGGCTTCAACAGTGCATCGCA 1394
644 ThrArgIleAspArgProSerIleHisSerGln 654
|||||
1395 ACCGAGATTGACAGGCCATTCATCACAGCCAG 1427

seq_name: /sids1/gcgdata/geneseq/geneseqn-emb1/NM2000.DAT:AAA27475
seq_documentation_block:
ID AAA27475 standard; DNA: 49136 BP.
XX
AC AAA27475;
XX
DE 15-AUG-2000 (first entry)
XX
DE NIDDM1 region including CAPN10 and GPR35 genes.
XX
KW NIDDM1; non-insulin-dependent diabetes mellitus; CAPN10 gene;
KW calpain 10; GPR35 gene; G protein coupled receptor; human;
KW polymorphism; UCSNP-43; chromosome 2; diaphragm-1; diagnosis;
XX therapy; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT 1235..1515
FT /*tag= a
FT /*number= 1
FT /*codon_start= 1375
FT 1516..3812
FT /*tag= b
FT /*number= 1
FT 3813..3944
FT /*tag= c
FT /*number= 2
FT 3945..5282
FT /*tag= d
FT /*number= 2
FT 5283..5479
FT /*tag= e
FT /*number= 3
FT 5283..5468
FT /*tag= f
FT /*number= 3*

|||||
5480..6400
FT /*tag= g
FT /*number= 3
FT 5469..6400
FT /*tag= h
FT /*number= 3*
FT 6401..6618
FT /*tag= i
FT /*number= 4
FT 6619..8372
FT /*tag= j
FT /*number= 4
FT 8373..8514
FT /*tag= k
FT /*number= 5
FT 8515..9009
FT /*tag= l
FT /*number= 5
FT 9010..9175
FT /*tag= m
FT /*number= 6
FT /*note= TGA, 9013..9015*
FT 9176..9490
FT /*tag= n
FT /*number= 6
FT 9491..9771
FT /*tag= o
FT /*number= 7
FT 9772..10399
FT /*tag= p
FT /*number= 7
FT 10400..10618
FT /*tag= q
FT /*number= 8
FT /*note= TGA, 10455..10457*
FT 10619..10784
FT /*tag= r
FT /*number= 8
FT 10785..10987
FT /*tag= s
FT /*number= 9
FT 10988..11146
FT /*tag= t
FT /*number= 9
FT 11147..11408
FT /*tag= u
FT /*number= 10
FT 11316..11408
FT /*tag= v
FT /*number= 10*
FT 11409..12353
FT /*tag= w
FT /*number= 10
FT 12354..12553
FT /*tag= x
FT /*number= 11
FT /*note= TGA, 12412..12414*
FT 12554..12817
FT /*tag= y
FT /*number= 11
FT 12818..12863
FT /*tag= z
FT 12864..13116
FT /*tag= aa
FT /*number= 12
FT 13117..13569
FT /*tag= ab
FT /*number= 13
FT /*note= TAA, 13144..13146*
FT 13570..30856
FT /*tag= ac
FT /*number= 13

```

CC polypeptides and nucleic acids, and using these modulators to treat
CC diabetes, in particular through the regulation of an insulin
CC secretory response or insulin mediated glucose transport.
XX
SQ Sequence 49136 BP; 10300 A; 13118 C; 14165 G; 11553 T; 0 other;

alignment_scores:
Quality: 1989.00 Length: 3020
Ratio: 3.224 Gaps: 11
Percent Similarity: 20.430 Percent Identity: 20.298

alignment_block:
US-09-768-877-2 x AAA27475

Align seg 1/1 to: AAA27475 from: 1 to: 49136

47 GlnGluileCysAlaThrProArgLeuPheProAspProArgGluG1 63
|||||
3810 CAGGAGATTGTGCCACACCCCGCTGTTCCAGATGACCCAGGGAAG 3859
63 yGlnVallyGlnGlyLeuLeuGlyAspCysTrpPheLeuCysAlaCysA 80
|||||
3860 GCAGGTTAAAGCAGGGGCTGCTGGGGATTGCTGTTCTCTGTGCTGCG 3909
80 laAlaLeuGlnLysSerArgHisLeuLeuAsp..... 90
|||||
3910 CCGCGCTGCAGAGACGACGACCTCTCGACAGAGTGGGGGCCCTTTC 3959
90 90

3960 CCTGTGTTGTCCTCGAGCGGGTTCCTTTTGGTTTCTCCAGCCTGCTG 4009
90 90

4010 AGTACCAGGAGGCTTTCGAAAGCAGAGCTGTGCCGACGCCGATCTCT 4059
90 90

4060 GCTGTGTTGGGGAAGGAGGAGAGAGTTCACAGGACAGAGCTGAGGACTGC 4109
90 90

4110 ACTCTGTCCCTCTGTCGAGGGGGGGTGGCTTGGCTGCCAGAGAGGCTTC 4159
90 90

4160 CATCAGGAGGTTGCGCCCTGCTCTGTGCTCTCTGACCCCGGACTCCAT 4209
90 90

4210 GGAGTCAGATCACCAGGTTTGAATAAAGAGACAAATGTGCCAGCTCACA 4259
90 90

4260 GGAGGAGGGGCTGCTGCGACGCTCTGCTCAGATCTCTCTCAGCTAG 4309
90 90

4310 CTCGCTGTTTCTTACAGGTTTGAATAAAGTTCGAAAAAGTTATTA 4359
90 90

4360 AACCTGTTTCTGTGGGTAGACAGATATCTTGGGAGGAGAGGCTTCTCA 4409
90 90

4410 GGTTCCTTACCTGGGAGTGTTCACCGTTTATGCTTGGCTTGTGCTA 4459
90 90

4460 AGTGTGCTGATTATGACGCGGCTCAACAGTGTGACCTTATTTCAGAGT 4509

FT exon 30857..30980
/*tag= ad
/number= 14
30981..31445
/*tag= ae
/number= 14
31446..32175
/*tag= af
/number= 15
replace(6225,a)
/*tag= ag
/frequency= 0.75
/note= "G-allele has a frequency of 0.75 in
Mexican Americans"
/frequency= 0.71
/note= "G-allele has a frequency of 0.71 in
non-Hispanic whites of German ancestry"
/frequency= 0.90
/note= "G-allele has a frequency of 0.90 in
African Americans"
/frequency= 0.94
/note= "G-allele has a frequency of 0.94 in
Asians (Japanese)"
43070..44944
/*tag= ah
/note= "encodes GPR35 (AA79576)"
44922..44927
/*tag= ai

PN WO200023603-A2.
XX 27-APR-2000.
PD
XX
XX 21-OCT-1999; 99WO-US24890.
XX
PR 21-OCT-1998; 98US-0105052.
PR 13-MAY-1999; 99US-0134175.
XX
PA (ARCH-) ARCH DEV CORP.
XX
XX Polonsky KS, Horikawa Y, Oda N, Sreenan S, Zhou Y, Otani K;
PI Hanis CL, Bell GI;
XX
XX WPI; 2000-339702/29.
DR P-PSDB; RAY79567, RAY79568, RAY79569, RAY79570, RAY79571, RAY79572,
DR RAY79573, RAY79574, RAY79576.
XX
PT Method for screening for type 2 diabetes mellitus comprises detecting a
PT polymorphism in a calpain encoding nucleic acid segment or a
PT protease-encoding nucleic acid segment -
XX
PS Claim 65; Page 203-217; 257pp; English.
XX
XX The present sequence is that of a 49,136 bp region located within
XX the NIDDM1 region of human chromosome 2. It includes the CAPN10
XX gene that encodes a novel calpain-like cysteine protease, designated
XX calpain 10, and a gene encoding a G protein coupled receptor,
XX GPR35. Alternative splicing of calpain 10 mRNA generates a family
XX of proteins. Isoforms 10a-h (see RAY79567-74) are respectively
XX encoded by exons 1-7,9-13, 1-7,9,10*,11-13, 1-7,9,11-13,
XX 1-10*,11-13, 1-3*,4-7,9-13, 1,2,14,15 and 1,11-13. Calpain 10 mRNA
XX is ubiquitously expressed; the major 2.7 kb transcript was detected
XX in every human adult and foetal tissue examined. Mutations in the
XX CAPN10 gene are responsible for susceptibility to type 2 diabetes.
XX The nucleotide variant showing all the evidence for linkage to
XX type 2 diabetes, UCSNP-43, is located at nucleotide 6225 in intron
XX 3 of the gene. There is alternative splicing of intron 3, but the
XX molecular mechanism by which the polymorphism at UCSNP-43 affects
XX susceptibility to type 2 diabetes is unclear. Claimed methods for
XX screening for a propensity for type 2 diabetes mellitus are based
XX on detection of a polymorphism in a calpain encoding nucleic acid,
XX especially UCSNP-10 of the CAPN10 gene. Methods are also claimed
XX for identifying modulators of calpain activity using calpain 10

```
90 ..... 90
4510 TTCACATATGCCAGGCCCATGTAAAGCTGCACAGTCACTGGCTT 4559
90 ..... 90
4560 CAGACATATGCTTACCACTTGTACTTTTAAAAAGAGACTGCTA 4609
90 ..... 90
4610 AAATCACTGCCATTGTGTTCTGGCCGTAAGTGCACAGACGACGACGC 4659
90 ..... 90
4660 ACCAGTCCCGTGCCAGCCTGCACCTCCCGATCGGTGAGAAATGC 4709
90 ..... 90
4710 TTACCCCTGAAGAGCCCTCGGTGAGACCTGTGGGTCAAGACAGTGGC 4759
90 ..... 90
4760 CCAGAGTCGGGGCGCAGAGCTCCAGCAGCAGAGAGGGGCTGCTGCTCG 4809
90 ..... 90
4810 TGTGAGCTGTGCTTGACGCCAGCTCGTCAAGAACTGGGTCACTCATG 4859
90 ..... 90
4860 CCCTGATATGCATTTGTTGTTGGCTTCAGGTCAATGCTTTAGTAGAG 4909
90 ..... 90
4910 GGCAGCAGAGTGTGCCCGGATATGTGGCTCCTCGGTGTGCTCTCA 4959
90 ..... 90
4960 GTTTTGAATGAGAGGTCTGTTAATTCATGTGGTATGACGCCCTGTG 5009
90 ..... 90
5010 CAGCGCGCATCATCCAGTGTGCCGTAGAGTCTCTCGACATCCAGGTG 5059
90 ..... 90
5060 TGCCGTAGAGTCTCTGCGACATCCAGGTGTGCCGTAGAGTCTCTGCGA 5109
90 ..... 90
5110 CATCCAGGTGTCCGTAGAGTCTCTGCGACATCCAGGTGTGCGTTAGAG 5159
90 ..... 90
5160 TTCTCTGCAGACCGCGGTGCTGTGAGCACTACGCTGTGCCACACCGC 5209
90 ..... 90
5210 GGCCGGACACATGAGTACGCGCGGTGTATATACAGGCTCGGCTT 5259
91 ..... 91
5260 TTGCTTCTCCCTGTGATGCGAGTCAATCTCCGCGACAGCCAGAGTGG 5309
101 AlaSpIngluTyrArgLysPheThrCysArgIleTrrpGlnPheG1 117
|||||
5310 GCCGACCAAGAGTACCGGGGCTCTTCACTGTGCGCATTTGGCATTTGG 5359
117 YArgTrrpValGluValThrThrAspAspArgLeuProCysLeuAlaG1 134
|||||
5360 ACGCTGGGTGAGGTGACCAAGATGACCGCTGCGCTTGCAGAGGA 5409
134 rglLeuCysPheSerArgCysGlnArgGlnAspValPheTrrpLeuProLeu 150

|||||
5410 GACTCTGTTTCCCGCTGCCAGAGGAGAGATGTCTTGTGCTCCCTTA 5459
151 LeuGlnLysValTyrAlaLys..... 157
5460 CTGMAAAGGTCTACGCCCAAGTGCCTGTGTGGGGCTGAAGGGCTGGC 5509
157 ..... 157
5510 CTGGGGCAAGTGGAGCTCCACTACATAGGCTGCCAGAGAGTCTC 5559
157 ..... 157
5560 TGTCACTCTGGGCTGCAGAGCCCTTCACTTCTGAGAGTCTGCACT 5609
157 ..... 157
5610 CATTCGTGACTCAGGCTGACAGGCCAGGTGACAGATTCYTTTGGG 5659
157 ..... 157
5660 CTTGTGATTGCCACATCCCTGCTTTCCTTCCCTTCCAAAGCCAG 5709
157 ..... 157
5710 CGTGAGTCTGTTCCACAGAGAAATGTGTCCGCTCCTCTATTAT 5759
157 ..... 157
5760 CGGCCCAAGAAAGATGCTCTTATATTGTGTGAAATGTTGG 5809
157 ..... 157
5810 ACAGCAGACTCATTTGTGTAGTCTGTGGGAGAGTAGGCTACCCAGC 5859
157 ..... 157
5860 ATACCAACACTTGTATCAAGGTGCTTGTGGCTCAGGGGACCAAGACC 5909
157 ..... 157
5910 CTCACATGATCATTAATTAATAGCCTTCCCTCTAGAAATGATTTGTC 5959
157 ..... 157
5960 TTCTTCCAAAGCACTGACCTGACAGAGCAGGAGGAAGCTGTGAAC 6009
157 ..... 157
6010 ATGGAGAGCTGTGCTGTGACATCATGCGCAGTACCCCTTCCATCCA 6059
157 ..... 157
6060 AGGCTGTTTTAGAAAAGCAGGTTGAGCTTGAGAGCAAGGATGTG 6109
157 ..... 157
6110 GGCATCATATGCTTCCAGGCTCCTGCTGCTGTCTGTGCCACACCGGA 6159
157 ..... 157
6160 TGCCAGAGACTTCTGTGTGGCAGAGAGACTGACAGGGGCTACAGCTT 6209
157 ..... 157
6210 GCTGCAAGTAAGCGTTTGAAGTGTAGGCTTAAGCCTTGACTTGTGAGG 6259
157 ..... 157
6260 ATGAGGAAGAGCAGAGGGAGTAAAGAGTGGATTTAGGACAGCGTT 6309
157 ..... 157
```

6310 GGAGATTGGGGTGCTACAGACCATTGGGAATCAGAGAGGGGGCCATGCT 6359
158ValHisG1 160
6360 CAATGCCAGAGGCTCACTCCCATGGTGATTGTCTCCCTAGGTCCTCATGG 6409
160 ySerTyrGluHisLeuTrpAlaGlyGlnValalaAspalaLeuValAspL 177
6410 GTCTACGAGCACCTGTGGCGGGCGAGGTGGCGGATGCCCTGGTGGACC 6459
177 euThrGlyGlyLeuAlaGluArgTrpAsnLeuLysGlyValalaGlySer 193
6460 TGACCGCGCGCTGGCAGAAAGATGGAACCTGAAGGGCGTAGCAGGAAGC 6509
194 GlyGlyGlnGlnAspArgProGlyArgTrpGluHisArgThrCysArgG1 210
6510 GGAGGCCACGACGACAGGCCAGGCCGCTGGAGCAGACAGACTGTGCGCA 6559
210 nLeuLeuHisLeuLysAspGlnCysLeuLysSerCysValLeuSerP 227
6560 GCTGTCCACCTGAAGGACCAGTGTCTGATCAGCTGCTGCTGCTCAGCC 6609
227 roArgAlaGly 230
6610 CCAGAGCAGGTGAGGCACGTCGGCCAAACATGGGAGGCTGCAGCCAGCGTG 6659
230 230
6660 CCCCCCACTGCCAGGCCCTCAGGCACACTGTAGCTTTTATGTACTGGCT 6709
230 230
6710 ACACAGCCCTGTACGACTTAAGTGGGAAGAAGTAAGCTTGTTCTCAAGGG 6759
230 230
6760 TGGTGTCTCAGTTGTGTGACCTTCCCTGCTGCTCTCTTCCAGAGGAGC 6809
230 230
6810 TGGCCCTTCTCTCCCTGACCACTCTTCCACTAGTCGAGGCGCAGGAAG 6859
230 230
6860 AGGTGGCACCGAGTCAAGGCCCACTGTCTGTGCCATCCCTGGCCCACTG 6909
230 230
6910 GCAACTGGCAAAATCAAAACCTGTTTTTTATTTAGTGATAGATAACAT 6959
230 230
6960 TCGTTAAAAACAGTTTGTCTCCAAAAAANTGAAAGGGCGCAGGTGTGGTG 7009
230 230
7010 CTCACACCTGTAAATCCAGAACTTTGGGAGGCTGAGTGGGAGTATCGCT 7059
230 230
7060 GGAGCCAGGAGTTCAAGAGACCCCGCTGGGCAACATGGCAAGATCTCAT 7109
230 230
7110 CTCACAAAAAANTGAAGAAAAAANTCACTTAGATGGAACACATGTGAC 7159
230 230
7160 TTTTGTAGTGCCTCTCAGTTTTTCCATGAGCAGCAGCGTTTACGTGTCC 7209
230 230
7210 CTTCCGACCGCTTCTCACATGCCACACAGCGCTGTCAAAATGTCCGC 7259

230 230
7260 CCATGAGCGGGTTTCCACAGTCTCTTAATCATTCCTCGATGTGGGATA 7309
230 230
7310 TTAAGATCTCTCTGTTTTATATATATATAAGCAGTATCAATGAACATCT 7359
230 230
7360 TTATCATTTTTTTCATACTTAGGATTAATTTAAAAATATGGTTAAAGAAT 7409
230 230
7410 ATGAATATCACAGTAAGTAACAAAGTTGTATAGGCCCTTGGTCTGGGT 7459
230 230
7460 CCCCCAATAAGACAGCCCTGAGATGAGTTTCAGGACACGTTGAGAGGTTTC 7509
230 230
7510 AGAGAGCCTGGAGGGCGTGTACTCCCGCCAGCCCCGCTGGGGTGCCAGG 7559
230 230
7560 AAGGACGCTGGCCTCAGAGCCTCCACCTGGGAGTGAAGGGCGCTGCTGT 7609
230 230
7610 GTGGSCATTAGCTCGGGGAGCTGTTGGTTTTGGGTGCTCCAGGGTGGGGT 7659
230 230
7660 AGCGCTAAGTCTTAGCACTTCAGGCTTTAGAGAAAGCCCCCAGGCAGAGA 7709
230 230
7710 GAGATGGCAGCTGGCAGTACTGGAGTGCATTTGGAGCCTGCTGAGGTGG 7759
230 230
7760 CAAGGGCGCGGCTGGCGGAGACTGAAGTCCCTTCCAGAGAGACCTCCT 7809
230 230
7810 CTCTAGACCGTTGCCCGCCACAGATGTGGCTTCTCTCTGTTTGGGCTC 7859
230 230
7860 TTTCTTTTCTATGCCCATCTGTCTCAAGGTCTCTGGCCAGGTGGAGCTGT 7909
230 230
7910 TGCCCCCTGGGCTCCCATGCTCCAGTGGGCTTCTAGCCAGGCGACTAT 7959
230 230
7960 CTGAATGTCTTACCTCTAGCAGGCTTTGCCAAGAGAAACACTCTCTAGAT 8009
230 230
8010 TTCTGGCAGTTGCAGATGCATCTGTGCAGCATGTTTGAGAACTGGTGTCTG 8059
230 230
8060 TGCCAGGCATCGTCACAGATGGGTACGGAGATGACTGAGGCGCTAGAGA 8109
230 230
8110 AGGGGATGACTTACCAGCACAGACCCGGATGGCAGCCGAGTCAGGCC 8159

9060	CCGAGCTCCAGGAAGGAGGAGTTC	TGGGCTGGAGGAGAGAGTTC	TCCAGG	9109
311	GIuPhaspGIuLeuThyAlaGlyTrp	ProValIThrGluAlaGlyHisLe	327	
9110	GAGTTTGACAGCTCCACCGTTGGCTA	ACC	CGGTCACGGAGGCGCGC	CACCT
327	uGIuSerLeuTrpThr			332
9159	GCAGACCTCTTACACAGAGTAGTAC	CCCCGAGGGGCTGTGCTGGGACGTGC		9206
332			332
9209	TCThCTGCCGAGTAGAGGAGCTGGG	GACGGTGCTGGATTCCCTTCG		9258
332			332
9259	CCAGGCCCAAGTTGGTCTCTTCA	CGTGGAGAGATGATTCTGTCCAGG		9308
332			332
9309	AGCCGGGAGGAGGGTATGATTCTG	TGCCAGAGCTGGGAGGAGGGTGG		9358
332			332
9359	CTTTGGGGAGGGGCTGGCTCTGT	CTGTGGCGGTAGCTGCTGTTAGACC		9408
332			332
9409	TGCCAGGTTTCATGAGGCCACCGT	GGGCGGAGCCAGGAGAGGCCGTGT		9458
333	GIuArgLeuLeuGlyHisThr		339
9459	CCCACAGCTGATCCCTGGTGTTC	TCTCACTAGAGGCTGCTGCCATA		9508
339	hTrArgAlaLeuProGlyAlaTrp	ValIysGlyGlnSerAlaGlyLys		355
9509	CGGGGGCGCTGGCTGGGGCTGGG	CTCAAGGGCCAGTCAACATGAAGCTGC		9558
356	ArgAsnAsnSerGlyPheProSer	AsnProLysPheTrpLeuArgValSe		372
9559	CGGAACACAGCGGCTTTC	CCAGCAACCCAAATCTGGCTGGGGTTC		9608
372	rGIuProSerGIuValTrpIleAla	ValLeuGlnArgSerArgLeuHisA		389
9609	AGAACCGAGTGAAGGTACATTG	CCGCTCGCAGAGATCCAGCGTCACG		9658
389	IaAlaAspTrpAlaGlyArgAla	ArgAlaLeuValGlyAspSerHisThr		405
9659	CGGGGACACGGGACAGCGCGG	CGCGCCGACCTGGTGGTACAGTCACTACT		9708
406	SetTrpSerProAlaSerIlePro	GlyLysHisTrpGlnAlaValGlyLe		422
9709	TGCTGGAGGCCACGACAGATCC	CGGGCAAGCATACACAGCTGTGGGTCT		9758
422	uHISLeuTrpLysVal	GIu		428
9759	GCACCTCTGGAAAGTAGTCA	ACCCCCGCTGGCTCACGCTCGTTCA		9808
428			428
9809	GGTGGTGTGAGAGCCCATG	AGAGTCTGGGTTCTAGAGATGGCTCTG		9858
428			428
9859	GACACATGTGACTCTGCC	ACGGCCCCACACAGTCTCCCCCTCTTG		9908
428			428
9909	TGTTGCACGGGGTTGAC	CTTGTGTGTCTCTCCACAGCCGGGCTGTG		9958
428			428

9959 GAGACTGCAGCTCTTCTGCTTCCCGTGTGCTCATTTGGCCAAAGGAAA 10008
428 428
10009 GCACAGAGTCTGCAGCCAGGACCGCAGGAGGCGCTGGACCCG 10058
428 428
10059 GGGGCTCTTGGCAGCGCGTCTTCTGAGCAAGAGGTAGAGCCAG 10108
428 428
10109 CGGCTGAGGACCTGTACAGGCCAGTCCAGCTGTGCAGTGTGTGTGA 10158
428 428
10159 CCTGGCACACATCTCTCCCTGCTCCCTCAGTCTTTCCTCCCTGCAGAC 10208
428 428
10209 GGGTCTGTACAGGATCTCATGGATTGCTGTGAGGCCAGGAGTCCC 10258
428 428
10259 AGGCTAACCACTGGTTCAAAAGTGTGTTTCCAGGAAGAACAGATG 10308
428 428
10309 GGGGCGCTCAGGGCAAGGCGCTGAGTGTGTCGAGGATATCCCGGCTG 10358
428 428
10359 CTCGCTCAGGGGCTGGGTTTTCATCTGTGTGTGTTGACAGGTGTGACA 10408
428 428
10409 CTTGGCACCACTGTTCCTGTCCCTTCATGATGTGCCCCACATGATG 10458
428 428
10459 TTCCTTCTCTTCAAAAGAGTGTGTGAAGGCCCACTGTCCAGCAGC 10508
428 428
10509 CCCCAGGTTGCCTGGGCACAGTGTGCTTGTGGGCCCACTACAGGAGG 10558
428 428
10559 ACTTGCAGGCTCGTCTGGGACAGATCTGGCGCCAGGGCCAAAGTGAAG 10608
428 428
10609 CCGGGATTGGTGGGCATCTCTAGTGTGCTCCCTGAGAGAGGTGGAGGT 10658
428 428
10659 GCTGACAGGCTTGGCGCTTTCATGTCAACTCCAGAGGCCCTGTGTCT 10708
428 428
10709 TGCAGCAGGAGGTCAAGGCCAGGCGCTGTGACCCCGCGCTCTCCAC 10758
429
10759 ACTGAGCTCTGCAGTGTCTACAGGTAGAGAAGCGCGGTCAATCTG 10808
435 ProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAlaTy 451
10809 CTTAGGGTCTCTGTCCATGCCCGCGGTGGTGGCACCGGTGTCATGCATA 10858
451 rAspArgGluValHisLeuArgCysGlnLeuSerProGlyTyrTyrLeuA 468
10859 CGACCGGAGGTCCACCTCGGTTGTGAGCTCTCACCGGCTACTACCTGG 10908

468 laValProSerThrPheLeuLysAspAlaProGlyGluPheLeuLeuArg 484
10909 CTCTCCCAAGCAGCTTCTCTGAAGCAGCGCCAGGGAGTCTCTCTCCGA 10958
485 ValPheSerThrGlyArgValSerLeu..... 493
10959 GTCTTCTCTACCGGCGAGTCTCTCTTAGGTAGAGGAACCGCCAGTGC 11008
493 493
11009 TGCTGGCTCTCCGAGGACACAGGCCCTTCCAAGGCAGGATTGGGCACCT 11058
493 493
11059 TCCTCTGTGGTGGCAGGTGTCCATGTGGAACTGAGGCCACTTGGGAAC 11108
494 SerAlaIleArgAl 498
11109 CTGCTGCCAGCGCCCTCCCATGTCTTCTTCTGGCAGCGCCATCAGGGC 11158
498 aValAlaLysAsnThrThrProGlyAlaAlaLeuProAlaGlyGluTrpG 515
11159 AGTGGCAAGACACACACCGCGGCGAGCTGCTGCGGGGAGTGGG 11208
515 lyThrValGlnLeuArgGlySerTrpArgValGlyGlnThrAlaGlyGly 531
11209 GGACCTGTGAGCTACCGGGTCTTGGAGAGTGGCCAGACGGCGGGGGC 11258
532 SerArgAsnPheAlaSerTyrProThrAsnProCysPheProPheSerVa 548
11259 AGCAGGAACCTTGGCTCATACCCACCAACCCCTGCTTCCCTCTCGGT 11308
548 lProGluGlyProGlyProArgCysValArgIleThrLeuHisGlnHisC 565
11309 CCGCAGGCGCTGGCGCGCTGCTGCCATCACTCTGTCATCAGCACT 11358
565 ysArgProSerAspThrGluPheHisProIleGlyPheHisIlePhe... 580
11359 GCGGCGGAGTGACACCGAGTTCACCCCATCGCTTCCATATCTTCCAG 11408
580 580
11409 GCAAGCTCCTTGCCTCCAGGAGGGGGGAGGAGGAGGCGCTCAGAG 11458
580 580
11459 AATTTCATCTTGGCTTCCATTGTCCCAACAGAGGCTCTGGGCTCAGTC 11508
580 580
11509 ACTTGGGCTCCCTGCTCCCTTCGAGCGCTGCTAGAACCCGACAGGGC 11558
580 580
11559 CCTCTCCCATCTCCAACCTCTCAGAGGCAAGGCCGAAGATGGCTCTGGA 11608
580 580
11609 AGGGCGGGGGCTTGGAGGTGGGCGAGGCTGATCCAGGCAGGCGAGT 11658
580 580
11659 TCCAGAGGAGTGGTGTAGTGGGAGGAGGAGAGTGGAGAGGACAG 11708
580 580
11709 GAGGCGGAGTGTGAGACAGCGGGGTGGTGGCGCTTGGCTTGGGAAC 11758
580 580
11759 GCAGGGGCTGTGACTGAGAGTCAAGAGTGAAGAGGAGGAGGCCAGGCTGGC 11808

```

580 ..... 580
11809 TTGGCCACAGACCCCTGGGTGTGAGAGATCCACAGTCACTGACTCA 11858
580 ..... 580
11859 GCCACAGACCCCTGTCCACTTACCTGACTGAGAAATGACTGTCCAG 11908
580 ..... 580
11909 GTTCAATTCTGCAGACATGTGTCCCTGGAATGAGGGGCCCTGACGAG 11958
580 ..... 580
11959 GAAGGACACTGACACCTCGGTTTACAGTGGGCTGGGGAGCCCTTGA 12008
580 ..... 580
12009 CCCTGCTGTTTGGCTTGGGCGACCGGCTCAGTCCCTAGAGCTTGAG 12058
580 ..... 580
12059 GAAACACATGCCAGGCGCAGTGGAGCCCTTGGGGGGGGGCTGGGCACTC 12108
580 ..... 580
12109 ACAAGTGTAAAGCCCTGATGATGAGACAGCCCTCAGGCGGGGCCCC 12158
580 ..... 580
12159 ACTGCCGGCACCTTCTGGCAAGGCTGGCCAGGCTTGGTGAGAGGCGAG 12208
580 ..... 580
12209 TCCAGTGTCCAGGCTGGAGCCCTCTCAGAGAAAGGGGCTGATGTGA 12258
580 ..... 580
12259 CTCAGAGGGCCAAAGGCGATCCGAGCAGATGGCTGGGCTCCCT 12308
581 ..... 583
12309 ACCCCAGGCTGGCCCCCTCAGTCTGAGCCCTTCTCAGGCTCC 12358
583 ..... 600
12359 AGAGGGTGAAGAGCAGAGCAGCAGCAGCTGCTCTCAGAGAGCGC 12408
600 ..... 616
12409 TGGTGAAGTGTGTCACATGCTAGCCAGAGGTGAGCCGCTGTC 12458
617 ..... 633
12459 CTCCTGCTGGGGCACCTTACAGGTTGCTCCCTCAGCTACCTGCCGA 12508
633 ..... 647
12509 CACAGAGGGGCTTCCACAGTGCATCCGACACAGAGATTGACAGGGG 12558
647 ..... 647
12559 GCTGTGGAGCTTGGGGGGCGGCGAGCTGAGAGCTGGGTCTGAGTCTTA 12608
647 ..... 647
12609 GTGCTGCTGTCCTCCACAGCTCTCTGCTGCCCTCAGCCCTCAAGCC 12658
647 ..... 647
12659 CTATCTGCTGTGGACAGCAGGCTCTCTGCTTACCTGAGGAGCCCTTCC 12708
647 ..... 647

```

```

12709 TTGCTGTGAGGCTGAGAGAGAGTCTAGTGGAGGTGGGCCAGGAGC 12758
647 ..... 647
12759 ACACAGCCACTTGTGTGACAAAGTCACTGGAGCGCTGATGTGTC 12808
648 ..... 662
12809 TCTCCACAGGCGCATTCATTCACAGCAGAGATGCTGGCCAGTCC 12858
662 ..... 664
12859 AAGAGGTG 12866

seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH04003

seq_documentation_block:
ID   AAH04003 standard; cDNA; 874 BP.
XX
AC   AAH04003;
XX
DT   26-JUN-2001 (first entry)
XX
DE   Human cDNA clone (5'-primer) SEQ ID NO:838.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
EP   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999; 99JP-0248036.
XX
PR   27-AUG-1999; 99JP-0300253.
XX
PR   11-JAN-2000; 2000JP-0118776.
XX
PR   02-MAY-2000; 2000JP-0183767.
XX
PR   09-JUN-2000; 2000JP-0241899.
XX
XX
XX   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX   WPI; 2001-318749/34.
XX
PT   primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs.
XX
PS   Claim 1; SEQ ID 838; 2537pp + CD ROM; English.
XX
XX
XX   The present invention describes primer sets for synthesizing 5602
XX   full-length cDNAs defined in the specification. Where a primer set
XX   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX   to the complementary strand of a polynucleotide which comprises one of
XX   the 5602 nucleotide sequences defined in the specification, where the
XX   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX   of an oligonucleotide comprising a sequence complementary to the
XX   complementary strand of a polynucleotide which comprises a 5'-end
XX   sequence and an oligonucleotide comprising a sequence complementary to a
XX   polynucleotide which comprises a 3'-end sequence, where the
XX   5'-end sequence/3'-end sequence is selected from those defined in
XX   the specification. The primer sets can be used in antisense therapy and
XX   in gene therapy. The primers are useful for synthesizing polynucleotides,
XX   particularly full-length cDNAs. The primers are also useful for the
XX   detection and/or diagnosis of the abnormality of the proteins encoded by
XX   the full-length cDNAs. The primers allow obtaining of the full-length

```

CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 874 BP; 161 A; 238 C; 315 G; 155 T; 5 other;

alignment_scores:
Quality: 1251.00 Length: 286
Ratio: 4.849 Gaps: 7
Percent Similarity: 90.210 Percent Identity: 86.364

alignment_block:
US-09-768-877-2 x AAH04003

Align seg 1/1 to: AAH04003 from: 1 to: 874

145 valPheTrpLeuProLeuLeuGluLysValTyrAlaLysValHisGlySe 161
|||||
5 GTGTTCTGGCTCCCTTACTTGGAAAGAGTCTACGCCAAGGTCCATGGTGC 54
|||||
161 rTyrGluHisLeuTrpAlaGlyGlnValAla.AspAlaLeuValAspLeu 177
|||||
55 CTACGAGCACTGTGGCCCGGCGAGGTGGTGGATGCCCTGTGGACCTG 104
|||||
178 ThrGlyGlyLeuAlaGluArgTrpAsnLeuLysGlyValAlaGlySerG1 194
|||||
105 ACCGGCGGCTGSCCAAGATGTAACCTGAAGGGCGTAGCAGGAAGCGG 154
|||||
194 yGluGlnAspArgProGlyArgTrpGluHisArgThrCysArgGlnL 211
|||||
155 AGCCCAAGCAGGACAGACAGCGCGCTGGGAGCACAGGACTTGTGGCAGC 204
|||||
211 euLeuHisLeuLysAspGlnCysLeuLysSerCysCysValLeuSerPro 227
|||||
205 TGTCTCACTGAGGACAGTGTCTGATCAGCTGCTGCGTGTACGCCCC 254
|||||
228 ArgAlaGlyAlaArgGluLeuGlyGluPheHisAlaPheIleValSerAs 244
|||||
255 AGAGCAGGTGCCCGGAGCTGGGGAGTTCATGCTTCCATGCTCTCGGA 304
|||||
244 pLeuArgGluLeuGlnGlyGlnAlaGlyGlnCysIleLeuLeuLeuArgI 261
|||||
305 CTGCGGGAGCTCCAGGGTCCAGCGGGCCAGTGCATCTCTGCTGCGGA 354
|||||
261 leGlnAsnProTrpGlyArgGlyCysTrpGlnGlyLeuTrpArgGluGly 277
|||||
355 TCCAGAACCCCTGGGGCCCGCGGTGCTGGCAGGGGCTCTGGAGAGAGGG 404
|||||
278 GlyGluGlyTrpSerGlnValAspAlaValAlaSerGluLeuLeuSe 294
|||||
405 GGTGAAGGCTGGAGCCAGGTAGATCAGCGGTAGCATCTGAGCTCTGTC 454
|||||
294 rGlnLeuGlnGluGlyGluPheTrpValGluGluGluPheLeuArgG 311
|||||
455 CCAGCTCCAGGAAGGGAGTCTTGGTGGAGGAGGAGTCTCTCAGGG 504
|||||
311 luPheAspGluLeuThrValGlyTyrProValThrGluAlaGlyHisLeu 327
|||||
505 AGTTTGACGAGCTACCCGTGGTACCCGGTACCGAGGCGCCGACCTG 554
|||||
328 GlnSerLeuTyrThrGluArgLeuLeuCysHisThrArgAlaLeuProG1 344
|||||
555 CAGAGCCTACACAGAGAGGTGCTGTGCCATACGCGGGCGCTGCTGG 604
|||||
344 yAlaTrpValLysGlyGlnSerAlaGlyCysArgAsnAsnSerGlyP 361
|||||
605 GGCCTGGGTCAAGGGCCAGTACGAGGAGGTGTCGCGGAACAACAGCGCT 654
|||||
361 heProSerAsnProLysPhe.TrpLeuArgValSerGluProSerGluVa 377
|| |||||

655 TTNCAGCAACCCCAAAATCTTGTTCGGGTCTCAGAACCCGAATGAGGT 704
377 lTyrIleAla...ValLeuGlnArgSerArgLeu...HisAlaAlaAspT 392
|||||
705 GTACATTTGGCCGCTCTGTGCAAGATCCAGGCTTGAACGCCGCGGACT 754
392 rp.....AlaGlyArgAlaArgAlaLeuValGly...AspSerHisThr 405
|| |||||
755 GGGCAAGGCGGGCGGCGGCGGCACTTGGTGGGTGACAGTCTATTACTT 804
406 SerTrpSerPro...AlaSerIleProGlyLysHisTyrGlnAlaValG1 421
|||||
805 TCTGGAAGCCCAACGAGGCACTTCCCGGGCGNAGCACTTACCAAGCTTGG 854
421 yLeu 422
|||
855 GGTT 858

seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AAH99362

seq_documentation_block:

ID AAH99362 standard; cDNA; 751 BP.

XX AAH99362;

DT 16-OCT-2001 (first entry)

XX Human protein encoding cDNA sequence SEQ ID NO:197.

XX Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
KW anti-inflammatory; antirheumatic; antiarthritic; immunosuppressive;
KW antibacterial; endocrine; cardiant; central nervous system; virucide;
KW anti-HIV; fungicide; antimutagen; cardiovascular; antianemic; anaemia;
KW antiagregant; haemostatic; antiallergic; antidiabetic; osteopathic; eczema;
KW dermatological; antiallergic; antidiabetic; antidiabetic; infection;
KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
KW immunostimulant; gene therapy; antitense therapy; vaccine; inflammation;
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
KW thrombocytopenia; osteoporosis; severe combined immunodeficiency;
KW allergic rhinitis; diabetes; multiple sclerosis; depression;
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
KW neurological disorder; ss.

OS Homo sapiens.

XX WO200153455-A2.

XX 26-JUL-2001.

XX 22-DEC-2000; 2000WO-US35017.

XX 23-DEC-1999; 99US-0471275.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-457603/49.

DR P-PSDB; AAM25421.

XX Isolated human polynucleotides encoding polypeptides, useful for the treatment and diagnosis of e.g. cancer, ulcers and HIV infection -

Claim 1; Page 386; 1217pp; English.

XX AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
CC AAM25963. The proteins can have activities based on the tissues and
CC cells they are expressed in, such as: antiinflammatory; antirheumatic;
CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;

CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;
CC cardiovascular; antianemic; antiagregant; haemostatic; vulnerary;
CC antifuror; osteopathic; dermatological; antiallergic; antiasthmatic;
CC antidiabetic; cytostatic; neuroprotective; antidepressant; nocotropic;
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC encoding them can be used in gene therapy, antisense therapy and vaccine
CC production. The proteins and polynucleotides are useful for screening for
CC agonists or antagonists of a protein and for the treatment and diagnosis
CC of disorders associated with the activity of a protein e.g. inflammation,
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC anaemia, platelet disorders, thrombocytopenia, wounds, burns, ulcers,
CC osteoporosis, severe combined immunodeficiency, eczema, allergic
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC neurological disorders.
XX
SQ Sequence 751 BP; 133 A; 219 C; 255 G; 144 T; 0 other;

alignment_scores:
Quality: 964.00 Length: 195
Ratio: 5.446 Gaps: 1
Percent Similarity: 90.769 Percent Identity: 89.744

alignment_block:

US-09-768-877-2 x AAH99362 ..

Align seg 1/1 to: AAH99362 from: 1 to: 751

60 ProArGInGInGInValLysGInGLeuLeuGInGAspCysTrpPhe 76
|||||
2 CCACGGGAAGGGCAGGTAGACAGGGGCTGCTGGGGATGCTGCTCT 51
76 uCysAlaCysAlaAlaLeuGInLysSerHisLeuLeuAspGInVal 93
|||||
52 GTGTGCTGCGCGCGCTGCAGAAAGACGACCTCTGCGACAGTCA 101
93 LeProGInGInProSerTrpAlaAspGInGLeuTrpArgLysSerPhe 109
|||||
102 TTCTCCCGGACAGCCGAGCTGGCCGACACAGAGTACCGGGCTCTTC 151
110 ThrCysArgLLeuTrpGInPheGInLysArgTrpValGInValThr 126
|||||
152 ACCTGTCGATTTGGCAGTGGAGCTGGGTGGAGTGACACAGATGA 201
126 PArgLeuProCysLeuAlaGInArgLeuCysPheSerArgCysGInArg 143
|||||
202 CCGCTGCGCGCTGCTGCAGGGAGACTGTTCTCCCGCTGCCAGAGGG 251
143 LuAspValPheTrpLeuProLeuLeuGInLysValTyrAlaLysValHis 159
|||||
252 AGGATGTCTTGGCTCCCTTACTGGAAGGTCTACGCCAGGTCAT 301
160 GlyseryTrGInHisLeuTrpAlaGInGInValAlaAspAlaLeuValAs 176
|||||
302 GGGTCTCTAGACACCTGTGGCCGGCAGGTGGGGATGCCCTGCTGA 351
176 PleuThrGInGInLeuAlaGInArgTrpAsnLeuLysGInValAlaGIn 193
|||||
352 CTTACCGCGCGCTGCGCAAAAGATGMACTGAAAGGCGCTACACAGAA 401
193 erGInGInGInAspArgProGInArgTrpGInHisArgThrCysArg 209
|||||
402 GCGAGGCGCAGACAGACAGCCGCTGGAGCACACAGGACTTGTGG 451
210 GInLeuLeuHisLeuLysAspGInCysLeuLLeuSerCysValLeuSe 226
|||||
452 CAGCTGCTCCACCTGAAGACCACTGTGATCAGCTGCTGCTGCTGAG 501
226 rProArgAlaGInValArgGInLeuGInGInPheHisAlaPheIleValS 243
|||||

502 CCCACAGCAGGT 514
243 erAspLeuArgGInLeuGInGInGInAlaGIn 254
|||
515GAGGCACGTGGCCACAGCAGGAGG 538

OM of: US-09-768-877-2 to: Issued_Patents_NA.* out_format : pfs

Date: Sep 17, 2002 6:34 AM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL=itame+g2n.model -DEV=xlp
-/cg2n2_1/USPTOPOOL/US09768877/runat.16092002.152812.8369/app-query.fasta.1.739
-BA-Issued_Patents_NA -OFMT=fasta -SUFFIX=tri -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000
-FCGAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FCGAPOP=6.000 -FCGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blsum62
-TRANS-human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pcr
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09768877_ECGNL_1.98 -NCPD=6 -ICPD=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1
```

Search information block:

```
Query: US-09-768-877-2
Query length: 672
Database: Issued_Patents_NA.*
Database sequences: 383533
Database length: 122816752
Search time (sec): 84.830000
```

score_list:

Sequence	Strid	Orig	ZScore	EScore	len	Documentation	
/cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-09-422-869-3				3649.00	6014.72	0.0	2620
/cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-09-422-869-13				3604.00	5940.88	0.0	2455
/cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-09-422-869-5				3290.00	5420.79	3.0e-294	2297
/cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-09-422-869-11				3229.50	5333.87	1.6e-289	2516
/cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-09-422-869-19				3090.50	5090.34	7.6e-264	2204
/cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-09-422-869-7				2963.00	4877.19	5.7e-264	2511
/cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-09-422-869-1				2983.00	4624.17	6.7e-172	49136
/cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-08-835-099A-5				635.00	1017.26	5.6e-49	2109
/cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-08-835-099A-6				635.00	1017.10	5.6e-49	2136
/cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-09-157-349-6				633.00	1017.10	5.8e-49	2136
/cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-09-422-869-15				513.50	822.07	4.2e-38	1267
/cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-09-422-869-17				472.00	758.00	1.6e-34	864
/cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-08-928-692-16				263.00	390.11	4.8e-14	4700
/cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-08-658-136-2				164.00	195.46	0.0034	53526
/cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-08-658-136-1				164.00	195.45	0.0034	53577
/cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-08-943-731-5				151.00	186.15	0.0111	20084
/cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-09-048-804-1				147.50	203.08	0.0013	3257
/cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-09-056-105-26				136.50	180.86	0.0219	4473
/cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-08-229-515A-9				136.50	180.86	0.0219	4473
/cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-08-645-865-9				136.50	180.71	0.0223	4530
/cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-09-335-409-1				136.00	145.88	1.94	68750
/cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-09-568-102-1				136.00	145.88	1.94	68750
/cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-09-568-969-1				136.00	145.88	1.94	68750
/cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-09-568-480-1				136.00	145.88	1.94	68750
/cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-09-568-486-1				136.00	145.88	1.94	68750
/cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-09-568-472-1				136.00	145.88	1.94	68750
/cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-09-103-840A-1				134.50	91.38	2.1e-03	4411529
/cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-08-476-866-20				132.50	164.07	0.1883	10079
/cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-08-804-227C-7				132.50	145.55	2.03	44377
/cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-08-804-198-1				132.50	145.55	2.03	44377
/cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-08-188-582-31				132.00	175.11	0.0457	3901
/cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-08-646-715-31				132.00	175.11	0.0457	3901
/cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-09-150-460B-3				131.00	168.51	0.1066	6614
/cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-09-443-501A-2				131.00	137.01	6.06	71989
/cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-09-422-869-21				130.50	181.78	0.0194	1875
/cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-08-306-691B-42				130.00	164.46	0.1792	7011
/cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-09-103-840A-2				129.00	82.28	6.8e+03	4403765

```
/cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-09-404-650-1 + 128.50 162.33 0.2356 681
/cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-09-404-650-3 + 128.50 162.26 0.2378 685
/cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-08-123-161A-13 + 128.00 174.60 0.0488 2
/cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-08-483-278-13 + 128.00 174.60 0.0488 23
/cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-08-068-945A-1 + 127.00 153.27 0.7531 11
```

seq_name: /cg2n2_6/ptodata/2/lna/6B.COMB.seq:US-09-422-869-3

```
seq_documentation_block:
Sequence 3, Application US/09422869
Patent No. 6235481
GENERAL INFORMATION:
APPLICANT: POLONSKY, KENNETH S.
APPLICANT: HORIKAWA, YUKIO
APPLICANT: ODA, NAOHISA
APPLICANT: COX, NANCY J.
APPLICANT: SREENAN, SEAMUS
APPLICANT: ZHOU, YUN-PING
APPLICANT: OTANI, KENICHI
APPLICANT: HANIS, CRAIG L.
APPLICANT: BELL, GRAEME I.
TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
FILE REFERENCE: ARCD:307
CURRENT APPLICATION NUMBER: US/09/422,869
CURRENT FILING DATE: 1999-10-21
EARLIER APPLICATION NUMBER: 60/134,175
EARLIER FILING DATE: 1999-05-13
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 2620
TYPE: DNA
ORGANISM: Human
US-09-422-869-3
```

```
alignment_scores:
Quality: 3649.00 Length: 672
Ratio: 5.430 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
```

alignment_block:
US-09-768-877-2 x US-09-422-869-3 ..

Align seq 1/1 to: US-09-422-869-3 from: 1 to: 2620

```
1 MetAAGAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
|||||
178 ATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 227
|||||
17 aaAlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrPro 34
|||||
228 CCGCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 277
|||||
34 euAlaGluPheArgGluAspLeuThrThrPArgArgProGluGluIleCys 50
|||||
278 TGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 327
|||||
51 AlaThrProArgLeuPheProAspAspProArgGluGluGluIleValysgl 67
|||||
328 GCCACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 377
|||||
67 ngLyLeuLeuGlyAspCysThrPheLeuCysAlaCysAlaAlaLeuGlu 84
|||||
378 GGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 427
|||||
84 ySerThrGHisLeuLeuAspGluValIleProProGluGluInProSerTrp 100
|||||
428 AGGACGACGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 477
|||||
101 AlaAspGluGluIleThrArgGlySerPheThrCysArgIleThrPheGlu 117
|||||
478 GCCGACGACGACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 527
|||||
```

117 yAGTrpValgluValThrThraspAspArgLeuProCysLeuAlaGlyA 134
528 ACCTGGTGGAGGTGACCACAGATGACCGCTGCGTGCCTTCGAGGGA 577
134 rgLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
578 GACTCTGTTTCTCCCGTCGCAGAGGAGGATGTCTTGGCTCCCTTTA 627
151 LeuGluLysValTyrAlaLysValHisGlySerTyrGluHisLeuTrpAl 167
628 CTGGAAAGGCTTACGCCAAGGTTCATGGGTCCCTACGAGCACCTGTGGG 677
167 aglyGlnValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGluA 184
678 CGGCGAGGTGGCGATGCCCTGGTGGACCTGACCGCGCGCTTGGCAGAA 727
184 rgTrpAsnLeuLysGlyValAlaGlySerGlyGlyGlnGlnAspArgPro 200
728 GATGGAAACCTGAAGGCGGTAGCAGGAAGCGGAGGCCAGCAGACAGGCCA 777
201 GlyArgTrpGluHisArgThrCysArgGlnLeuLeuHisLeuLysAspGl 217
778 GCGCGCTGGGAGCACAGGACTTGTCCGAGAGTGTCTCCACCTGAAGGACCA 827
217 nCysLeuIleSerCysValLeuSerProArgAlaGlyAlaArgGluL 234
828 GTGCTGTATCAGCTGCTGGTGCTCAGCCCCAGACAGGTGCCGGGAGC 877
234 euGlyGluPheHisAlaPheIleValSerAspLeuArgGluLeuGlnGly 250
878 TGGGGAGGTTCATGCTTCATTTGCTCGGACCTGCGGGAGCTCCAGGCT 927
251 GlnAlaGlyGlnCysIleLeuLeuLeuArgIleGlnAsnProTrpGlyAr 267
928 CAGCGGGGCGATGTGATCTGCTGTCGGGATCCAGAACCCCTGGGCGC 977
267 gArgCysTrpGlnGlyLeuTrpArgGluGlyGlyGluGlyTrpSerGlnV 284
978 GCGGTGCTGSCAGGGGCTCTGGAGAGAGGGGGGTGAAGGTGTGAGCCAGG 1027
284 alaSpAlaIaValAlaSerGluLeuLeuSerGlnLeuGlnGluGlyGlu 300
1028 TAGATGCACGGTAGCATCTGAGTCTCTGCCAGCTCCAGGAAGGGGAG 1077
301 PheTrpValGluGluGluPheLeuArgGluPheAspGluLeuThrVa 317
1078 TTTCTGGTGGAGGAGGAGGTTCTCTCAGGGAGTTTACGAGCTCACCGT 1127
317 lGlyTrpProValThrGluAlaGlyHisLeuGlnSerLeuTyrThrGluA 334
1128 TGGCTACCGGTTACGGAGGCGGCCACCTGCAGAGAGCTCTACACAGAGA 1177
334 rgLeuLeuCysHisThrArgAlaLeuProGlyAlaIleTrpValLysGlyGln 350
1178 GGCTGCTCTGCATACGCGGGCGCTGCTGCGGGCTGGGTCAAGGGCCAG 1227
351 SerAlaGlyGlyCysArgAsnAsnSerGlyPheProSerAsnProLysPh 367
1228 TCACGAGGAGGCTGCCGGAAACAACAGCGGGTTTCCACAGCAACCCCAAT 1277
367 eTrpLeuArgValSerGluProSerGluValTyrIleAlaValLeuGlnA 384
1278 CTGGCTGGGGTCTCAGAACCGAGTGAGGTGTACATTTGCCCTCTCCAGA 1327
384 rgSerArgLeuHisAlaAlaAspTrpAlaGlyArgAlaArgAlaLeuVal 400
1328 GATCAGAGCTGCACGCGCGGACTGGGACGCGCGGCGCGGCGGCGCTGG 1377
401 GlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHisTy 417
1378 GGTGACAGTCATACTCTGTGGAGCCCAAGCAGAGATCCCGGGCAAGCACTA 1427

417 rGlnAlaValGlyLeuHisLeuTrpLysValGluLysArgValAsnL 434
 1428 CCAGGCTGTGGGTCTGCACCTCTGGAAGGTAGAGAAGCGCGGTCAATC 1477
 434 euProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAla 450
 1478 TGCCTAGGGTCTGTCCATGCCGCCCGGTGGCTGSCACCGGTGCCATGCA 1527
 451 TyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTyrTyrLe 467
 1528 TAGCACC GGAGGTCCACCTTGGTGTGAGCTCTCACCGGGCTACTACT 1577
 467 uAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeuLeuA 484
 1578 GGGTGTCCCCAGCACCTTCTGAAGGACGCCACGGGGAGTCTCTGTCTCC 1627
 484 rGValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAla 500
 1628 GAGTCTTCTTACCGGGCGAGTCTCCCTTAGCGCCCATCAGGGCAGTGGCC 1677
 501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyGluTrpGlyThrVa 517
 1678 AAGAACACACCCCGGGGAGCCCTGCTTCCGGGGAGTGGGGGACCGT 1727
 517 lGlnLeuArgGlySerTrpArgValGlyGlnThrAlaGlyGlySerArgA 534
 1728 GCAGCTACGGGGTCTTGGAGAGTGGCCAGACGCCGGGGGGCAGCAGGA 1777
 534 snPheAlaSerTyrProThrAsnProCysPheProPheSerValProGlu 550
 1778 ACTTTGGCTCATACCCACCAACCCCTGCTTCCCTTCTCGGTCCCGCAG 1827
 551 GlyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArgPr 567
 1828 GGCCCTGGCCCCGGTGGCTCCGCATCACTCTGCATCAGCACTGCCGGCC 1877
 567 oSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValProG 584
 1878 CAGTGACCGGAGTTCACGCCCATCGCTTCCATATCTTCCAGGTCCCGAG 1927
 584 luGlyGlyArgSerGlnAspAlaProProLeuLeuGlnGluProLeu 600
 1928 AGGTTGGAAGGAGCAGGACGACGCCGCCCTGCTGCTGCAGGAGCGCTG 1977
 601 LeuSerCysValProHisArgTyrAlaGlnGluValSerArgLeuCyste 617
 1978 CTGAGTGGTGGCCACATCGCTAGGCCCAGAGGTGAGCGGCTCTGCTCT 2027
 617 uLeuProAlaGlyThrTyrLysValValProSerThrTyrLeuProAspT 634
 2028 CCTGCTGCAGGCACCTACAAGGTTGTGGCTTCCACCTTACCTGCCGACA 2077
 634 hrGluGlyAlaPheThrValThrIleAlaThrArgIleAspArgProSer 650
 2078 CAGAGGGGGCCCTTCACAGTGCACNTCGCAACAGGATTTGACAGGCCATCC 2127
 651 IleHisSerGlnGluMetLeuGlyGlnPheLeuGlnGluValSerValMe 667
 2128 ATTACAGCCAGGAGTCTGGGCCAGTCTCTCCAGAGGTTCTCCGTCTAT 2177
 667 tAlaValMetLysThr 672
 2178 GGCAGTGTGATAAACCC 2193

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-422-869-13

seq_documentation_block:

; Sequence 13, Application US/09422869

; Patent No. 6235481

GENERAL INFORMATION:

APPLICANT: POLONSKY

; APPLICANT: HORIKAWA, YUKIO

APPLICANT: ODA, NAOHISA

/ IN CONTACT: ODA, NAUGHTSA

APPLICANT: COX, NANCY J.
APPLICANT: SREENAN, SEAMUS
APPLICANT: ZHOU, YUN-PING
APPLICANT: OTANI, KENICHI
APPLICANT: HANIS, CRAIG L.
APPLICANT: BELL, GRAEME I.
TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
FILE REFERENCE: ARCD:307
CURRENT APPLICATION NUMBER: US/09/422,869
EARLIER FILING DATE: 1999-10-21
EARLIER APPLICATION NUMBER: 60/134,175
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 13
LENGTH: 2455
TYPE: DNA
ORGANISM: Human
US-09-422-869-13

alignment_scores:
Quality: 3604.00 Length: 673
Ratio: 5.395 Gaps: 2
Percent Similarity: 99.257 Percent Identity: 99.257

alignment_block:

US-09-768-877-2 x US-09-422-869-13 ..

Align seg 1/1 to: US-09-422-869-13 from: 1 to: 2455

1 MetArgLaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
|||||
42 ATGGGGCGGGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 91
17 AlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrProL 34
|||||
92 CGCCTTCGCGCGCGGAGCTCCTGCTCTTGAGCTGCTACGCGCGC 141
34 euaIaGlnPheArgGluAspIleThrTrpArgArgProGlnGluIleCys 50
|||||
142 TGGCCCGGTCGCGGAGACATCATGTGAGGCGCGCGCGCGCGGAGATTGT 191
51 AlaThrProArgLeuPheProAspAspProArgGluGlnValLysGln 67
|||||
192 GCCACACCGCGCGCTTTCCAGATGACCCAGGAGGAGGAGGAGGAGCA 241
67 ngIyLeuLeuGlyAspCysTrpPheLeuCysAlaCysAlaAlaLeuGln 84
|||||
242 GGGCTGTGCTGGGGGATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 291
84 ySerArgHisLeuLeuAspGlnValIleProProGlnGlnProSerTrp 100
|||||
292 AGAGCAGGCGACCTCTGACAGCATCTCTCGGACAGCGAGAGTGG 341
101 AlaAspGlnGluTrpArgGlySerPheThrCysArgIleTrpIlePheG 117
|||||
342 GCCGACACGAGATACCGGGGCTCTTACCTGTGCAATTTGGCAGTTGG 391
117 yArgTrpValGluValThrThrAspAspArgLeuProCysLeuAlaGly 134
|||||
392 AGCGTGGTGGAGTGACACAGATACCGCTGCGGCTGCTTGGCAGGA 441
134 rGluCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
|||||
442 GACTGTGTTCTCCGCTGCGCAGAGGAGATGTGTTGCGCCCTTA 491
151 LeuGluLysValTrpAlaLysValHisGlySerTrpGluHisLeuTrp 167
|||||
492 CTGGAA.....AAGGCTCATGTGCTTACGAGACACTGTGGG 529
167 LaGlyGlnValAlaAspAlaLeuValAspLeuThrGlyLeuAlaGlu 183
|||||

530 CCGGCGAGGTGGCGGATGCCCTGTGTGACCTGACCGCGCGCTGGCAGAA 579
184 ArgTrpAsnLeuLysGlyValAlaIleSerGlyGlyGlnAlaAspArgP 200
|||||
580 AATGGAACCTGAAGCGGTACAGAGAACCGAGGCGCAGACAGAGCGCC 629
200 oGlyArgTrpGlnHisArgThrCysArgGlnLeuLeuHisLeuLysAsp 217
|||||
630 AGCGCGCTGGGAGACACAGACTTGTGCGAGCTGCTCCACCTGAAGAGCC 679
217 IncysLeuIleSerCysCysValLeuSerProArgAlaGlyAlaArgGlu 233
|||||
680 AGTGTCTATACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 729
234 LeuGlyGluPheHisAlaPheIleValSerAspLeuArgGluLeuGlnG 250
|||||
730 CTGGGGAGTTCCATGCTTCTATGCTTCTGCGAGCTGCGGAGCTCCAGAG 779
250 yGlnAlaGlyIncysIleLeuLeuLeuArgIleGlnAsnProTrpGly 267
|||||
780 TCAGCGGGCCAGTGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 829
267 rArgCysTrpGlnGlyLeuTrpArgGluGlyGlyGlyTrpSerGln 283
|||||
830 GCGGTGCTGCGAGGGGCTGTGAGAGAGGGGCTGAAGGGTGGAGCCAG 879
284 ValAspAlaAlaValAlaSerGluLeuLeuSerGlnLeuGlnGlyGly 300
|||||
880 GTATATGACGCGGTAGCATCTGACCTCTGCTGCTGCTGCTGCTGCTGCT 929
300 uPheTrpValGlnGluGluGluPheLeuArgGluPheAspGluLeuThr 317
|||||
930 GTCTGTGGTGAGAGAGAGGAGTTCTCTGAGGAGTTTACAGACTTACCG 979
317 aLgIyTrpProValThrGluAlaGlyHisLeuGlnSerLeuTrpThrGlu 333
|||||
980 TTGGCTACCGCGTCCAGGAGCGCGCCACCTGCAGAGCTCTTACACAGAG 1029
334 ArgLeuLeuCysHisThrArgAlaLeuProGlyAlaTrpValLysGly 350
|||||
1030 AGGTGCTCTGCTCCATACCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 1079
350 nSerAlaGlyLysArgAsnAsnSerGlyPheProSerAsnProLysP 367
|||||
1080 GTGAGCAGGAGGCGCGCGGAACACACAGCGCTTCCACAGCAACCCAAAT 1129
367 heTrpLeuArgValSerGluProSerGluValTrpIleAlaValLeuGln 383
|||||
1130 TCTGGCTGCGGCTCTCAGAACCCAGTAGAGTGATGATGCTGCTGCTGCT 1179
384 ArgSerArgLeuHisAlaAlaAspTrpAlaGlyArgAlaArgAlaLeu 400
|||||
1180 AGATCCAGGCTGACCGGCGGAGCTGGGAGGCGCGGCGCGGCGACTGT 1229
400 lGlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHis 417
|||||
1230 GGCTGACAGTCACTTCTGCGAGCCCGACGACATCTCCGGGAAGGACT 1279
417 yGlnAlaValGlyLeuHisLeuTrpLysValGlnLysArgAlaValAsn 433
|||||
1280 ACCAGGCTGTGGGTGTGCACTCTGGAAGGTAAAGAAAGCGCGGCTCAAT 1329
434 LeuProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAl 450
|||||
1330 CTGCTAGGCTCTCTGCAATGCCCCGCTGCTGCTGCTGCTGCTGCTGCT 1379
450 aTyraAspArgGluValHisLeuArgCysGlnLeuSerProGlyTrpTyrl 467
|||||
1380 ATACGACCGGAGGTCCACCTCGTTGTGAGCTTCCACCGGGCTACTACCC 1429
467 euaIaValProSerThrPheLeuLysAspAlaProGlyGluPheLeu 483
|||||
1430 TGGCTGTCCCGACACCTTCTGGAAGAGCGCGCAGGAGTTCTGCTC 1479

```

484 ArgValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAl 500
|||||
1480 CGAGTCTTCTACCGGGGAGTCTCCCTAGCGCATCAGGGCAGTGGC 1529
|||||
500 alyAsnThrThrProGlyAlaAlaLeuProAlaGlyGluThrGlyThrV 517
|||||
1530 CAAGAACACACCCCGGGGAGCCCTGCTGGGGGAGTGGGGACCG 1579
|||||
517 aGlnLeuArgGlySerThrArgValGlyGlnThrAlaGlyGlySerArg 533
|||||
1580 TGCAGCTACGGGGTCTTGGAGAGTGGCCAGACGGGGGGGACAGG 1629
|||||
534 AsnPheAlaSerTyrProThrAsnProCysPheProPheSerValProG 550
|||||
1630 ACTTTGCTCTATACCCACCAACCCCTGCTTCCCTTCTCGGTCCCGA 1679
|||||
550 uGlyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArgP 567
|||||
1680 GGGCCTGGCCCGCTGGTCCGCATCACTCTGCATCAGCACTGCCGGC 1729
|||||
567 roSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValPro 583
|||||
1730 CCAGTGACACCGAGTCCACCCATCGGCTTCCATATCTTCCAGGTCCA 1779
|||||
584 GluGlyGlyArgSerGlnAspAlaProProLeuLeuGlnGluProle 600
|||||
1780 GAGGTGGAGGAGCCAGGACCCACCCCTGCTGCTGCAGGAGCCGCT 1829
|||||
600 uLeuSerCysValProHisArgTyrAlaGlnGluValSerArgLeuCysL 617
|||||
1830 GCTGAGTGGTGGCCACATCGCTACGCCAGGAGGTGAGCCGGCTCTGCC 1879
|||||
617 euLeuProIaGlyThrTyrLysValValProSerThrTyrLeuProAsp 633
|||||
1880 TCCTGCTCAGGACACCTACAAAGGTGTGCCCTCCACCTACCTGCCGGAC 1929
|||||
634 ThrGluGlyAlaPheThrValThrIleAlaThrArgIleAspArgProSe 650
|||||
1930 ACAGAGGGGGCTTACAGTGACCATCGCACCCAGGATGACAGGCCATC 1979
|||||
650 rIleHisSerGlnGluMetLeuGlyGlnPheLeuGlnGluValSerValM 667
|||||
1980 CATTCACAGCCAGGAGATGCTGGCCAGTCTCTCCAAAGAGGTCTCCGTCA 2029
|||||
667 etAlaValMetLysThr 672
|||||
2030 TGGCAGTGATGAAACC 2046

```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-422-869-5'

seq_documentation_block:

; Sequence 5', Application US/09422869

; Patent No. 6235481

; GENERAL INFORMATION:

; APPLICANT: POLONSKY, KENNETH S.

; APPLICANT: HORIKAWA, YUKIO

; APPLICANT: ODA, NAOHISA

; APPLICANT: COX, NANCY J.

; APPLICANT: SREENAN, SEAMUS

; APPLICANT: ZHOU, YUN-PING

; APPLICANT: OTANI, KENICHI

; APPLICANT: HANIS, CRAIG L.

; APPLICANT: BELL, GRAEME I.

; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES

; FILE REFERENCE: ARCD:307

; CURRENT APPLICATION NUMBER: US/09/422,869

; CURRENT FILING DATE: 1999-10-21

; EARLIER APPLICATION NUMBER: 60/134,175

; EARLIER FILING DATE: 1999-05-13

; NUMBER OF SEQ ID NOS: 30

; SOFTWARE: PatentIn ver. 2.0

; SEQ ID NO 5

; LENGTH: 2297

; TYPE: DNA

; ORGANISM: Human

US-09-422-869-5

alignment_scores:

Quality: 3290.00 Length: 672

Ratio: 5.350 Gaps: 1

Percent Similarity: 91.518 Percent Identity: 91.518

alignment_block:

US-09-768-877-2 x US-09-422-869-5

Align seg 1/1 to: US-09-422-869-5 from: 1 to: 2297

```

1 MetArgAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
|||||
42 ATGCGGGCGGGCGGGCGGACGCGCGGGAGGAGCTGTTCGGGAGCGC 91
|||||
17 aAlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrProL 34
|||||
92 CGCCTTCCCGCGCGGACTCTCTGCTCTTCTGCGACTTGTCTACGCCGC 141
|||||
34 euAlaGlnPheArgGluAspIleThrTrpArgArgProGlnGluIleCys 50
|||||
142 TGGCCCACTTCCGCGAGGACATCAGTGGAGCGGCCCGCCAGGAGATTGT 191
|||||
51 AlaThrProArgLeuPheProAspAspProArgGluGlyGlnValLysG 67
|||||
192 GCCACACCCCGGCTGTTTCCAGATGACCCACGGGAGGCGAGGTGAGCA 241
|||||
67 nGlyLeuLeuGlyAspCysTyrPheLeuLeuCysAlaCysAlaAlaLeuG 84
|||||
242 GGGGTGCTGGGGATTGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCT 291
|||||
84 ySerArgHisLeuLeuAspGlnValIleProProGlyGlnProSerTrp 100
|||||
292 AGAGCAGGCACCTCTCGACCAAGGTGATCTCCGGGACAGCCGAGCTGG 341
|||||
101 AlaAspGlnGluTyrArgGlySerPheThrCysArgIleTrpGlnPheG 117
|||||
342 GCCGACAGGAGTACCGGGGCTCTCTTCCCTGCTGCTGCTGCTGCTGCT 391
|||||
117 yArgTrpValGluValThrThrAspAspArgLeuProCysLeuAlaGly 134
|||||
392 ACCTGGGTGGAGGTGACACAGATGACCGCTGCTGCTGCTGCTGCTGCT 441
|||||
134 rgLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
|||||
442 GACTCTGTTTCTCCGCTGCCAGAGGAGGATGTTCTGCTGCTGCTGCTG 491
|||||
151 LeuGluLysValTyrAlaLysValHisGlySerTyrGluHisLeuTrpAl 167
|||||
492 CTGGAAGAGTCTACGCCAAGGTCCATGGGTCTTACGAGCACCTGTGGGC 541
|||||
167 aGlyGlnValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGlu 184
|||||
542 CGGCAGGTGGCGGATGCTGCTGGTGGACCTGACCGGGCGGCTGGCAGAA 591
|||||
184 rgtTrpAsnLeuLysGlyValAlaGlySerGlyGlyGlnGlnAspArgPro 200
|||||
592 GATGGAACCTGAAGGGCGTACAGAGACGGAGCGGCGGACGAGCAGGCCA 641
|||||
201 GlyArgTrpGluHisArgThrCysArgGlnLeuLeuHisLeuLysAspG 217
|||||
642 GGCGCTGGGAGCACAGGACTTGTGGCGAGCTGTCCACCTGAAGGACCA 691
|||||
217 nCysLeuIleSerCysValLeuSerProArgAlaGlyAlaArgGluL 234
|||||
692 GTGCTGTGATGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 741
|||||
234 euGlyGluPheHisAlaPheIleValSerLeuArgGluLeuGlnGly 250

```

```

|||||
742 TGGGGGAGTTCATGCTTCATGCTGAGACTGCGGAGAGCTCCAGGCT 791
251 GlnAlaGlyGlnCysIleLeuLeuLeuArgIleGlnAsnProTrpGlyAr 267
792 CAGCGGGGAGTGCATCTGCTGCTCGGATCCAGAAACCTCGGGCCG 841
267 GATGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 284
842 GCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 891
284 AlaAlaAlaAlaAlaAlaSerGluLeuLeuSerGlnLeuGlnGlyGlu 300
892 TAAATGACGCGGTAGCATCTGACCTCTGCTGCTGCTGCTGCTGCT 941
301 PheTrpValGluGluGluGluPheLeuArgGluPheAspGluLeuThrVa 317
942 TTTTGGGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 991
317 IglYTrpProValThrGluAlaGlyHisLeuGlnSerLeuTrpThrGlu 334
992 TGGCTACCGCGGTACGAGAGCGGCGGACCTGCGAGAGCTTACACAGA 1041
334 rglLeuLeuCysHisThrArgAlaLeuProGlyAlaTrpValGlyGln 350
1042 GGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1091
351 SerAlaGlyGlyCysArgAsnAsnSerGlyPheProSerAsnProLysPh 367
1092 TCAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1141
367 eTrpLeuArgValSerGluProSerGluValTrpIleAlaValLeuGln 384
1142 CTGGCTGCGGGCTCGAGAACCGAGTGTACATTCGCTGCTGCTGCT 1191
384 rGSerArgLeuHisAlaAlaAspTrpAlaGlyArgAlaAlaLeuVal 400
1192 GATCCAGGCTGCACCGCGGAGCTGGCGAGCGCGGCGGCGGCTGG 1241
401 GlYAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHisTy 417
1242 GGTGACAGTACTTCTGAGAGCCGAGGACATCCGCGGAGACACATA 1291
417 rGlnAlaValGlyLeuHisLeuTrpLysValGluLysArgArgValAsn 434
1292 CCAGGCTGTGGGTGCGACCTTGCAGAGTACAGAGAGAGAGAGAGAG 1341
434 eUrProAlaValLeuSerMetProProValAlaGlyThrAlaCysHisAla 450
1342 TGCCAGAGGCTCTGTCATGCCCGCGGTGGAGACCGCGGCGCATGCA 1391
451 TyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTrpTyrLe 467
1392 TACGACGCGGAGGTCCACCTGCTGTGAGCTTCACCGGCGCTACTACT 1441
467 uAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeuVal 484
1442 GCGTGTCCCGCAGACCTTCTGTAAGAGAGAGAGAGAGAGAGAGAGAG 1491
484 rGValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAla 500
1492 GAGTCTTCTTACCGCGGAGTCTCCCTTAG..... 1522
501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyLutTrpGlyThrVa 517
1522 ..... 1522
517 IglInLeuArgGlySerTrpArgValGlyGlnThrAlaGlyGlySerArg 534
1522 ..... 1522
534 snPheAlaSerTyrProThrAsnProCysPheProPheSerValProGlu 550

```

```

1522 ..... 1522
551 GlyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArgP 567
1523 GGCCTGTGCCCCCGCTGCGTCCGCTGCTGCTGCTGCTGCTGCTGCTG 1572
567 oSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValProG 584
1573 CAGTGCACCGAGTCCACCCCATCGGCTTCCATCTTCAGAGTCCAG 1622
584 LuGlyGlyArgSerGlnAspAlaProProLeuLeuGlnGluProLeu 600
1623 AGGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1672
601 LeuSerCysValProHisArgTyrAlaGlnGluValSerArgLeuCys 617
1673 CTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1722
617 uLeuProAlaGlyThrTyrLysValValProSerThrTyrLeuProAsp 634
1723 CTGCTGCGAGGACCTACAGAGTGTGCTGCTGCTGCTGCTGCTGCTG 1772
634 hrGlnGlyAlaPheThrValThrIleAlaThrArgIleAspArgProSer 650
1773 CAGAGGGGCGCTTACAGTGCATGCAACAGATGCAACAGATGCAAGCC 1822
651 IleHisSerGlnGluMetLeuGlnPheLeuGlnGluValSerValMet 667
1823 ATTCAACAGCCAGAGATGCTGGGCCAGATTCTCTCAAGAGAGTCTCC 1872
667 tAlaValMetLysThr 672
1873 GCGAGTGAATGAAGAAC 1888
seq_name: /cgn2_6/ptodata/2/1na/6B_COMB.seq:US-09-422-869-11
seq_documentation_block:
; Sequence 11, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: ODA, NAOHISA
; APPLICANT: COX, NANCY J.
; APPLICANT: SREENAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: HANIS, KENICHI
; APPLICANT: HANIS, CRAIG L.
; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
; FILE REFERENCE: ARCD:307
; CURRENT FILING DATE: 1999-10-21
; EARLIER APPLICATION NUMBER: 60/134,175
; EARLIER FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 2516
; TYPE: DNA
; ORGANISM: Human
US-09-422-869-11

```

alignment_scores:

```

Quality: 3239.50 Length: 745
Ratio: 5.267 Gaps: 2
Percent Similarity: 82.550 Percent Identity: 82.416

```

alignment_block:

```

US-09-768-877-2 x US-09-422-869-11 ..

```

```

Align seg 1/1 to: US-09-422-869-11 from: 1 to: 2516

```

1 MetArgAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
 42 ATCGGGCGGGCGGGCGGCGAGCGCGGAGGAGCTGTTCCGGGACGC 91
 17 aAlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrProL 34
 92 CGCCTTCCCGCGCGGGACTCTCGCTCTCTCGGACTTGTCTACGCCGC 141
 34 euAlaGlnPheArgGluAspIleThrTrpArgArgProGlnGluLeuCys 50
 142 TGGCCAGTTCGCGAGGACATCAGCTGGAGCGGCCCGCCAGGAGATTGT 191
 51 AlaThrProArgLeuPheProAspAspProArgGluGlnValIysG1 67
 192 GCCACACCCCGGCTGTTCCAGATGACCACCGGAAGGCGAGTGAAGCA 241
 67 nGlyLeuLeuGlyAspCysTrpPheLeuCysAlaCysAlaAlaLeuGlnL 84
 242 GGGGCTGCTGGGGGATGTGCTGTCTGTGCTGCTGCGCGCGCTGCAGA 291
 84 ysSerArgHisLeuLeuAspGlnValIleProProGlyGlnProSerTrp 100
 292 AGAGCAGGACCTCTCGGACCAAGTCAATCTCCGGGACAGCGAGCTGG 341
 101 AlaAspGlnGluTyArgGlySerPheThrCysArgIleTrpGlnPheG1 117
 342 GCGCACCAGGAGTACCGGGCTCTTCACTGTGCTGCTGCTGCTGCTGCT 391
 117 yArgTrpValGluValThrThrAspAspArgLeuProCysLeuAlaGlyA 134
 392 ACCTGGGTGGAGGTGACACAGATGACCGCTGCGCTGCTGCTGCTGCTG 441
 134 rgLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
 442 GACTCTGTTCTCCCGCTCCAGAGGAGGATGTGTTCTGGCTCCCTTA 491
 151 LeuGluLysValTyAlaLysValHisGlySerTyArgHisLeuTrpAl 167
 492 CTGGAAGAGTCTAGCCCAAGGTCTTCACTGCTGCTGCTGCTGCTGCT 541
 167 aGlyGlnValAlaAspAlaLeuValAspLeuThrGlyLeuAlaGluA 184
 542 CGGGCAGGTGGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 591
 184 rgTrpAsnLeuLysGlyValAlaGlySerGlyGlnGlnAspArgPro 200
 592 GATGAACCTGAAGGCGGTAGCAGAGAGGAGGAGCGGACGACAGGCGCA 641
 201 GlyArgTrpGluHisArgThrCysArgGlnLeuLeuHisLeuLysAspG1 217
 642 GCGCGCTGGGAGCACAGGACTTGTGGCAGCTGCTCCACCTGAAGGACCA 691
 217 nCysLeuIleSerCysValLeuSerProArgAlaGlyAlaArgGluL 234
 692 GTGTCTGATCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 741
 234 euGlyGluPheHisAlaPheIleValSerAspLeuArgGluLeuGlnGly 250
 742 TGGGGAGTTCATGCCITTCATGTCTCGGACTCTCGGGAGCTCCAGGGT 791
 251 GlnAlaGlyGlnCysIleLeuLeuLeuArgIleGlnAsnProTrpGlyAr 267
 792 CAGGGGGCGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 841
 267 gArgCysTrpGlnGlyLeuTrpArgGluGlyGlyGlyGlyGlyGlyGly 284
 842 GCGGTGCTGGCAGGGGCTCTGGAGAGAGGGGGTGAAGGGTGGAGCCAGG 891
 284 aAspAlaAlaValAlaSerGluLeuLeuSerGlnLeuGlnGluGlyGlu 300
 892 TAGATGCAGCGGTAGCATCTGAGCTCTCTCCAGCTCCAGGAGGGGAG 941

301 PheTrpValGluGluGluGluPheLeuArgGluPheAspGluLeuThrVa 317
 942 TTCTGGGTGGAGGAGGAGGAGTTCCTCAGGAGATTGTGAGGAGCTCACCGT 991
 317 lGlyTyProValThrGluAlaGlyHisLeuGlnSerLeuTyThrGluA 334
 992 TGGCTACCGCGTACGAGGAGCGCGCCACTGCAGAGCCTCTACACAGAGA 1041
 334 rGlyLeuLeuCysHisThrArgAlaLeuProGlyAlaTrpValLysGlyGln 350
 1042 GCTCTCTCTGCATACGCGGGCGCTGCTGCGGGCTGGGTCAAGGGCCAG 1091
 351 SerAlaGlyGlyCysArgAsnAsnSerGlyPheProSerAsnProLysPh 367
 1092 TCAGCAGGAGGCTGCCGGAACACAGCGGCTTCCACGACACCCCAANT 1141
 367 eTrpLeuArgValSerGluProSerGluValTyIleAlaValLeuGlnA 384
 1142 CTGGTGGGGTCTCAAAACCGAGTGAGGTGTACATTGCGCTCTGCAGA 1191
 384 rGSerArgLeuHisAlaAlaAspTrpAlaGlyArgAlaArgAlaLeuVal 400
 1192 GATCAGGCTGCACGCGGCGACTGGGCGAGCGCGGCGCGGCGACTGGTG 1241
 401 GlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHisTy 417
 1242 GGTGACAGTCACTCTCTGGAGCGCCAGGAGCATCCCGGGCAAGCACA 1291
 417 rGlnAlaValGlyLeuHisLeuTrpLys..... 426
 1292 CCAGCTGTGGGTCTGCACCTCTGGAAGGTGTGACACTTGGCACCCACAC 1341
 426 426
 1342 TGTTCCTCTCTCTTCATGGATGTGGCCACATGATGTTCCTCTCTT 1391
 426 426
 1392 GCAAAAGAAGTTGCTGGAAGGCCCACTCTCCAGCAGCCCGGAGTTGGCT 1441
 426 426
 1442 GGGCCACGGTGCCTTGTGGGCCAGCTTACAAGGAGGACTTGCAGGCTCG 1491
 427Val 427
 1492 TGTCTGGCAGAGTACTGCGCCAGGCGCAAGTGAAGCCCGGGATTGGTA 1541
 428 GluLysArgArgValAsnLeuProArgValLeuSerMetProProValAl 444
 1542 GAGAAGCGCGGTCAATCTGCTAGGCTCTGCTCCATGCCCCCGCTGGC 1591
 444 aGlyThrAlaCysHisAlaTyArgAspArgGluValHisLeuArgCysGluL 461
 1592 TGGCACCGCGTGCATGATGATACGACCGGAGGTCCACCTGCTGTGTGAGC 1641
 461 euSerProGlyTyTrpLeuAlaValProSerThrPheLeuLysAspAla 477
 1642 TCTCAGCGGGTACTACTGCTGCTGTCCACAGACCTCTCTGAAGGACGCG 1691
 478 ProGlyGluPheLeuLeuArgValPheSerThrGlyArgValSerLeuSe 494
 1692 CCAGGGGAGTTCTCTCCGAGTCTCTCTACCGGGCGAGTCTCCCTTAG 1740
 494 rAlaIleArgAlaValAlaLysAsnThrThrProGlyAlaAlaLeuProA 511
 1741 1741
 511 lAglyGluTrpGlyThrValGlnLeuArgGlySerTrpArgValGlyGln 527
 1741 1741
 528 ThrAlaGlyGlySerArgAsnPheAlaSerTyProThrAsnProCysPh 544

```

1741 ..... 1741
544 eproheservalProgluglyProargCysValargIlethrl 561
1742 .....GGCCCTGGCCCCCGCTGCTCCGATCATCTC 1772
561 euHISglHIScysArpProserAspThrIupheHisProIleIglyphe 577
1773 TGATTCACACACTGCCGCCAGTGACACGAGTTCCACCCATCGGCTTC 1822
578 HIsIlePheglInvalProgluglyGlyArgSerGlnAspAlaProProle 594
1823 CATACTTCCAGGTCCAGAGGGTGAAGAGCCAGCAGCACCCTCCACT 1872
594 uLeuEngIngluProLeuLeuSerCysValProHisArgTyrAlaIing 611
1873 GCTGCTGAGAGAGCCGCTGCTGAGCTGCTGCACATGCTACAGCTCCAGG 1922
611 luValSerArgLeuCysLeuLeuProAlaIglyThrTyrLysValValPro 627
1923 AGGTAGAGCGGCTGCTGCTGCTGCTGAGGACCTACAAAGTTGTGCCC 1972
628 SerThrTyrLeuProAspThrGluGlyAlaPheThrValThrIleAlaTh 644
1973 TCCACCTACCTGCGCGACACAGAGGGGCTTCACAGTACACATCGCMAAC 2022
644 fArgIleAspArgProSerIleHisSerGlnGluMetLeuGlyGlnPheI 661
2023 CAGATTTCACAGGCGATTCATTCACAGCAGAGAGATGCTGGGCCAGTTCC 2072
661 euGIngluValSerValMetAlaValMetLysThr 672
2073 TCCAAAGAGTCTCCGTCATGAGTGAATGAACCC 2107

```

seq_name: /cgn2_6/ptodata/2/lna/bb_COMB.seq:US-09-422-869-9

```

seq_documentation_block:
; Sequence 9, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: ODA, NAOHISA
; APPLICANT: COX, NANCY J.
; APPLICANT: SREENAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: OTANI, KENICHI
; APPLICANT: HANIS, CRAIG L.
; APPLICANT: BELL, GRAEME I.
; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
; FILE REFERENCE: ARCD:307
; CURRENT APPLICATION NUMBER: US/09/422,869
; EARLIER APPLICATION NUMBER: 60/134,175
; EARLIER FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 9
; LENGTH: 2204
; TYPE: DNA
; ORGANISM: Human
US-09-422-869-9

```

```

alignment_scores:
Quality: 3090.50      Length: 672
Ratio: 5.292          Gaps: 1
Percent Similarity: 86.905  Percent Identity: 86.905

```

alignment_block:
US-09-768-877-2 x US-09-422-869-9

Align seq 1/1 to: US-09-422-869-9 from: 1 to: 2204

```

1 MetArgAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
42 AVGGGGGGGGCGGGGGCGGACCCGGCGAGGAGCTGTTCCCGGAGCGC 91
17 aAlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrPro 34
92 GCGCTTCCCCCGCCGAGCTCTGCTCTTCTTGCGACTGTGTACGGCGC 141
34 euAlaGlnPheArgGluAspIleThrTrpArgArgProGlnGluIleCys 50
142 TGCCCGAGTTCCGGGAGACATCACTGAGAGCGGCCCCAGAGAAATTGT 191
51 AlaThrProArgLeuPheProAspAspProArgGluGlyValLysGly 67
192 GCCACACCCCGGCTGTTCCAGATGACCCAGGAGAGGCGCAGTGAAGCA 241
67 ngLYLeuLeuGlyAspCysTrpPheLeuCysAlaCysAlaAlaLeuGln 84
242 GGGGCTGCTGGGGGATTCGCTGCTGCTGCTGCTGCTGCTGCTGCA 291
84 ySeraArgHisLeuLeuAspGlnValIleProProGlyGlnProSerTrp 100
292 AGAGCAGGACACCTCTCGAGCAGCTCATTCCTCGGAGCAGCGAGCTGG 341
101 AlaAspGlnGluTyrArgGlySerPheThrCysArgIleTrpGlnPheG 117
342 GCCGACCAAGAGTACCGGGGCTCTTCACCTGCGCATTTGGCGATTGG 391
117 yArgTrpValGluValThrThrAspAspArgLeuProCysLeuAlaGly 134
392 ACCTGGGTGGAGGTGACACACAGATGACCGCTCCCTGCTTCACAGGA 441
134 rGlyLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuPro 150
442 GACTCTGTTTCTCCCGCTCCAGAGGAGATGTGTTCTGCTCCCTTA 491
151 LeuGluLysValTyrAlaLysValHisGlySerTyrGlnHisLeuTrpAl 167
492 CTGAAAGGCTACGCAAGGTCCATGGGCTCTACAGACACCTGTGGGC 541
167 agLYGlnValAlaAspAlaLeuValAspLeuThrGlyLYLeuAlaGlu 184
542 CGGGCAGTGGCGGATGCCCTGCTGAGCTGACCTGACGGGCTGCAAAA 591
184 rGTrpAsnLeuLysGlyValAlaGlySerGlyGlnGlnAspArgPro 200
592 GATGGAACCTGAAGGCGGTAGCAGAGAGCGAGGCCAGAGACAGCCA 641
201 GLYArgTrpGlnHisArgThrCysArgGlnLeuLeuHisLeuLysAspG 217
642 GGGCGTGGGAGCAGCAGACTTGTGGCAGCTGCTCCACCTGAAGGCCA 691
217 nCysLeuIleSerCysCysValLeuSerProArgAlaGlyAlaArgGlu 234
692 GTGCTCATACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 741
234 euGlyGluPheHisAlaPheIleValSerAspLeuArgGluLeuGlnGly 250
742 TGGGGAGATTCCATGCTTATTTGTCTGCGAGCTGGCGGAGCTCCAGG 791
251 GluAlaGlyGlnCysIleLeuLeuLeuArgIleGlnAsnProTrpGlyAr 267
792 CAGCGGGCCAGTCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 841
267 gArgCysTrpGlnGlyLeuTrpArgGluGlyGlyGlyGlyTrpSerGln 284
842 GCGGTGCTGGCAGGGGCTCTGAGAGAGAGGGGGTGAAGGGTGAAGCC 891
284 alaSpAlaAlaValAlaSerGluLeuLeuSerGlnLeuGlnGlyGlu 300
892 TAGATGACGGGTAGCATGTGAGCTCTGTCCTCCAGCTCCAGGAAGGGAG 941

```

```

301 pheTrpValGluGluGluPheLeuArgGluPheAspGluLeuThrVa 317
|||||
942 TTCTGGGTGAGGAGGAGGAGTTCTCAGGAGTTTGACGAGCTACCGT 991
|||||
317 lGlyTyrProValThrGluAlaGlyHisLeuGlnSerLeuTyrThrGluA 334
|||||
992 TGGCTACCCGGGTACAGGAGCGCGGCACCTGCAGAGCCCTCTACACAGA 1041
|||||
334 rGLeuLeuCysHisThrArgAlaLeuProGlyAlaTrpValLysGlyGln 350
|||||
1042 GCGTCTCTGCCATACCGCGGGCGCTGCCCTGGGGCGCTGGGTCAAGGCCAG 1091
|||||
351 SerAlaGlyGlyCysArgAsnSerGlyPheProSerAsnProLysPh 367
|||||
1092 TCAGCAGGAGGCTGCCGGAACAACAGCGCTTTCCAGCAACCCCAATT 1141
|||||
367 eTrpLeuArgValSerGluProSerGluValTyrIleAlaValLeuGlnA 384
|||||
1142 CTGGCTGCGGGTCTCAGAACCCAGTGAGGTGTACATTGCCGTCTGCAGA 1191
|||||
384 rGSerArgLeuHisAlaAlaAspTrpAlaGlyArgAlaArgAlaLeuVal 400
|||||
1192 GATCAGGCTGCACGCGCGGACTGGCAGCGCGCGCGCGGCACACTGGTG 1241
|||||
401 GlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHisTy 417
|||||
1242 GGTGACAGTCATACCTCGTGGAGCCAGCGAGCATCCCGGGCAAGCACTA 1291
|||||
417 rGlnAlaValGlyLeuHisLeuTrpLysValGluLysArgValAsnL 434
|||||
1292 CCAGGCTGTGGTCTGCACCTCTGGAAGGTAGAGAAGCGCGGGTCAATC 1341
|||||
434 euProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAla 450
|||||
1342 TGCTAGGCTCTGTCATGCCCCCGTGGTGGCAGCGCGGCCATGCA 1391
|||||
451 TyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTyrTyrLe 467
|||||
1392 TAGCAGCGGAGGTCCACCTGGTGTGAGCTCTACCGGGGTACTACCT 1441
|||||
467 uAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeuLeuA 484
|||||
1442 GGCTCTCCACAGCACCTCTCTGAAGGAGCGCGAGGGAGTTCTCTGCTCC 1491
|||||
484 rGValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAla 500
|||||
1492 GAGTCTTCTTACCGGGCGAGTCTCCCTAG..... 1522
|||||
501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyGluTrpGlyThrVa 517
|||||
1522 ..... 1522
|||||
517 lGlnLeuArgGlySerTrpArgValGlyGlnThrAlaGlyGlySerArgA 534
|||||
1522 ..... 1522
|||||
534 snPheAlaSerTyrProThrAsnProCysPheProPheSerValProGlu 550
|||||
1522 ..... 1522
|||||
551 GlyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArgPr 567
|||||
1522 ..... 1522
|||||
567 oSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValProG 584
|||||
1523 .....GTCCAG 1529
|||||
584 luGlyGlyArgSerGlnAspAlaProProLeuLeuGlnGluProLeu 600
|||||
1530 AGGTTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1579
|||||

```

```

601 LeuSerCysValProHisArgTyrAlaGlnGluValSerArgLeuCysLe 617
|||||
1580 CTGACTGGTGCACATCGCTACGCCAGGAGGTGAGCCGGCTCTGCCT 1629
|||||
617 uLeuProAlaGlyThrTyrLysValProSerThrTyrLeuProAspT 634
|||||
1630 CCTGCTGCAGGCACCTACAAGTTGTGCCCTCCACCTACCTGCCGACA 1679
|||||
634 hrGluGlyAlaPheThrValThrIleAlaThrArgIleAspArgProSer 650
|||||
1680 CAGAGGGGGCTTTCACAGTCACCATCGCAACACGAGATTGACAGGCATCC 1729
|||||
651 ileHisSerGlnGluMetLeuGlyGlnPheLeuGlnGluValSerValMe 667
|||||
1730 ATTACAGCCAGGAGATGCTGGCCAGTTCCTCCACAGAGGTCTCCGTAT 1779
|||||
667 talaValMetLysThr 672
|||||
1780 GGCAGTGATGAAACCC 1795
|||||

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-422-869-19

seq_documentation_block:
; Sequence 19, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: ODA, NAOHISA
; APPLICANT: COX, NANCY J.
; APPLICANT: SREENAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: OTANI, KENICHI
; APPLICANT: HANIS, CRAIG L.
; APPLICANT: BELL, GRAEME I.
; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
; FILE REFERENCE: ARCD:307
; CURRENT APPLICATION NUMBER: US/09/422,869
; CURRENT FILING DATE: 1999-10-21
; EARLIER APPLICATION NUMBER: 60/134,175
; EARLIER FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 2511
; TYPE: DNA
; ORGANISM: Mus musculus
; US-09-422-869-19

```

```

alignment_scores:
  Quality: 2963.00      Length: 671
  Ratio: 4.756         Gaps: 2
  Percent Similarity: 92.846  Percent Identity: 81.073

alignment_block:
US-09-768-877-2 x US-09-422-869-19 ..

Align seg 1/1 to: US-09-422-869-19 from: 1 to: 2511

```

```

1 MetArgAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
|||||
109 ATGGGGGGGTCCCGGGCGAGACCGCGCGGGGAGGAGCTTCCGGGAGCG 158
|||||
17 aAlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrProL 34
|||||
159 GGCATTCGCCCTCGGAGCTCTCGCTCTTTTACAACCTTGTCACGCGCTC 208
|||||
34 euAlaGlnPheArgGluAspIleThrTrpArgArgProGlnGluLeuCys 50
|||||
209 TGGCCCAAGTTTCGGGAGGAGATCATCTTGGAGACGACCCCGAGAAATCTGT 258
|||||
51 AlaThrProArgLeuPheProAspProArgGluGlyGlnValLysG1 67
|||||

```

```

|||||.....|
259 GCCACACCTCAGCTGTTTCCAGATTAACCCATGGGAGGACAGGTGAAGA 308
67 nglIleuLeuGlyAspCysTrpPheLeuCysAlaCysAlaLeuGln 84
|||||.....|
309 AAGGGCTGTGGAGATTGCTGGTCTGTGTGCTCTGCCGCCCTTCACA 358
84 ySerArghIstLeuLeuAspGlnValIleProProGlyInProSerTrp 100
|||||.....|
359 AAGTCAACACCTCCCTGGACAGGTCCTCCACAGACAGCAGGCTGG 408
101 AlaAspGlnGluTrpArgGlySerPheTrpCysArgIleTrpGlnPheI 117
|||||.....|
409 TGTGACAGAAATACCAAGGCTTCTTCACCTGTGGATTGGCAGATTGG 458
117 yArgTrpValGluValThrThrAspAspArgLeuProCysLeuAlaGly 134
|||||.....|
459 ACACGTGGGAGAGAGTACCATAGATGATGCTGGCTTGTCCGGGA 508
134 rglLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
|||||.....|
509 GACTGTGCTTTCCTCCGGTCCAGAGAGAGATGTTCTGGCTCCCTTA 558
151 LeuGlnCysValTrpAlaCysValHisGlySerTrpGlnHisLeuTrpAl 167
|||||.....|
559 CTGGAAGAGGCTATGCTTAAGTCCATGATGATGATGATGATGATGATG 608
167 aglyGlnValAlaAspAlaLeuValAspLeuThrGlyLeuAlaGln 184
|||||.....|
609 AGGGCAATGGCAGATGCTTAGTGATGATGATGATGATGATGATGATG 658
184 rgtTrpAsnLeuGlyValAlaGlySerGlyGlyGlnAspArgPro 200
|||||.....|
659 GTGGAGGCTTAAGATGATTAAGAAAGCCAGCCAGCAGACAGACCC 708
201 GlyArgTrpGlnHisArgThrCysArgGlnLeuLeuHisLeuLysAspG 217
|||||.....|
709 AGTGGTGGGAGACACAGAACTTGTGGCAGCTACTCCACTGAAGAGACG 758
217 nCysLeuIleSerCysValIleuSerProArgAlaGlyAlaArgGln 234
|||||.....|
759 GTGTCTATCAGCTGCTGTGCTTAGCCCAAGCAGAGCTGCCAGGGAC 808
234 euGlyGluPheHisAlaPheIleValSerAspLeuArgGlnLeuGlnGly 250
|||||.....|
809 TCGGAGAGTTCATGCTTCATCATCTCAGATCTGCAGAGCTCAGAGGT 858
251 GlnAlaGlyGlnCysIleLeuLeuLeuArgIleGlnAsnProTrpGlyArg 267
|||||.....|
859 CAGACTGGCCAGGGATATCTCTCTCTGCGGATTCACAAACCTCGGGCG 908
267 gArgCysTrpGlnGlyLeuTrpArgGlnGlyGlyGlnGlyTrpSerGln 284
|||||.....|
909 GCGTGTGGGAGGGCTCTGAGAGAAAGAGGTGAAGGGTGAACACAG 958
284 alaAspAlaIleValIleSerGlnLeuLeuSerGlnLeuGlnGlyGlu 300
|||||.....|
959 TAGAGCCAGCTAAAGAGTGTGAGTGTGCTGCCCAACTCCAGGAAGAG 1008
301 PheTrpValGlnGlnGluGluPheLeuArgGluPheAspGluLeuTrpVal 317
|||||.....|
1009 TTCTGGGTCCAGAGAGAGAGTCTCTCAGGAGTGTATGAGGTCAACCAT 1058
317 lGlyTrpProValThrGlnAlaGlnHisLeuGlnSerLeuTrpThrGln 334
|||||.....|
1059 CGGCTACCAAGTCAAGAGAGCCGGCCACACAGAGTCTCCACACAGAGA 1108
334 rglLeuLeuCysHisThrArgAlaLeuProGlyAlaTrpValLysGlyGln 350
|||||.....|
1109 GGGTGTGCTCCATACCGGACACTGCTGCTGCTGCTGCTGCTGCTGCT 1158
351 SerAlaGlyCysArgAsnAsnSerGlyPheProSerAsnProLysPhe 367
|||||.....|

```

```

1159 TCAGCAGAGAGCTGCCGGAACAACAGTTGCTTCCCTGCACACCCAGTT 1208
367 eTrpLeuArgValSerGluProSerGluValIleAlaValLeuGln 384
|||||.....|
1209 CTGGTTACGGCTGTGGAAACCCAGGAGAGTGTGTGCTGTGCTGTTCACA 1258
384 rGSerArgLeuHisAlaIleAspTrpAlaGlyArgAlaArgAlaLeuVal 400
|||||.....|
1259 GACCCCGGAGGGCTTA.....GTGGGCCAAGCTGGGCACTGGCG 1299
401 GlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHisTrp 417
|||||.....|
1300 GGT.....GCCAGTCCCTGACCGGGGAGACCTCCAGGAAAGACTA 1340
417 rGlnAlaValGlyLeuHisLeuTrpLysValGluLysArgArgValAsnL 434
|||||.....|
1341 CCAGGCTGTGGGCTGCACATCTGGAAAGTGAAGAAAGAAAGATCAGCC 1390
434 euProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAla 450
|||||.....|
1391 TGCCAGAGTCTCTGTCTGACCCCTGTGGCTGGCAGCTGCATGCGATGCG 1440
451 TyrAspArgGluValHisLeuArgCysGlnLeuSerProGlyTrpTrpLe 467
|||||.....|
1441 TATGATCGTGAGATCCACTGCTGTGTGAGCTCTCACCAGGCTACTACT 1490
467 uAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeuLeu 484
|||||.....|
1491 GGCGTCCCTGACACCTTTTGAAGATGTCCAGGGAGAGTTCCTGCTGCA 1540
484 rGValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAla 500
|||||.....|
1541 GAGCTTCTTACCTGAGGAATCTCCCTCAGTCCGCTGAGGCTGGCCACC 1590
501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyGluTrpGlyThrVal 517
|||||.....|
1591 AAGGTGCATGCCGTGMAACAGCCCTGCTCAGCGGAGTGGAGACTGT 1640
517 lGlnLeuArgGlySerTrpArgValGlyGlnThrAlaGlyLysArg 534
|||||.....|
1641 GCAGTGTGAGGGCTGTGAGAGAGCTGGCCACAGCTGGGGGAGCAGAGA 1690
534 snPheAlaSerTrpProThrAsnProCysPheProPheSerValProGln 550
|||||.....|
1691 ACTTGGCTGTACTCCCTGCAGATCCCTGCTCCCTTCTGTGTCTGTAG 1740
551 GlyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArgPr 567
|||||.....|
1741 GTGTCTGGCCCCGCTACATCCGATACACCTCAGCAAGCAACCTCGGCT 1790
567 oSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValProG 584
|||||.....|
1791 CAGTGACAGCCAGCTGCAACCCATGATGATGATGATGATGATGATGATG 1840
584 lueGlyGlyArgSerGlnAspAlaProProLeuLeuLeuGlnLysProLeu 600
|||||.....|
1841 CAGAGGCTGAGAACCAAGAGCGCGTTCCTGCTGCTGCTGCTGCTGCTG 1890
601 LeuSerCysValProHisArgTrpAlaGlnGlnValSerArgLeuCysLe 617
|||||.....|
1891 CTAAGCTGTGTACCATTCCTACGCCCAAGAGAGTGAAGCGCGCTGTGCT 1940
617 uLeuProAlaGlyThrTrpLysValValProSerThrTrpLeuProAspTr 634
|||||.....|
1941 CTTTCTGTGGGAACATACAGATTTGTTCCCTCACCCTCAGCTGCAGATA 1990
634 hrcLueGlyAlaPheThrValThrIleAlaThrArgIleAspArgProSer 650
|||||.....|
1991 CAGAGGCTACTCTCAGGTAACATGCAACCAAGAAATGATAGGCACTGCT 2040
651 lIleHisSerGlnGluLeuLeuGlyGlnPheLeuGlnGlnValSerValMe 667
|||||.....|
2041 ATCCACAGCAGAGAGATGCTGGGCCAGCTGCTCCAGAGGATCTCTTAT 2090

```


667 talavalMetLys 671
|||||
2091 GGCAGTCATGAAA 2103

seq_name: /cgn_2_6/ptodata/2/lna/6B_COMB.seq:US-09-422-869-7

seq_documentation_block:

; Sequence 7, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: ODA, NAOHISA
; APPLICANT: COX, NANCY J.
; APPLICANT: SREENAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: OTANI, KENICHI
; APPLICANT: HANIS, CRAIG L.
; APPLICANT: BELL, GRAEME I.
; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
; FILE REFERENCE: ARCD:307
; CURRENT APPLICATION NUMBER: US/09/422,869
; CURRENT FILING DATE: 1999-10-21
; EARLIER APPLICATION NUMBER: 60/134,175
; EARLIER FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 2001
; TYPE: DNA
; ORGANISM: Human
US-09-422-869-7

alignment_scores:

Quality: 2715.50 Length: 672
Ratio: 5.252 Gaps: 1
Percent Similarity: 76.935 Percent Identity: 76.786

alignment_block:

US-09-768-877-2 x US-09-422-869-7 ..

Align seg 1/1 to: US-09-422-869-7 from: 1 to: 2001

1 MetArgAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
|||||
42 ATGCGGGCGGGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 91
17 alaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrProL 34
|||||
92 CGCCTTCCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 141
34 euAlaGluPheArgGluAspIleThrTrpArgProGluGluIleCys 50
|||||
142 TGGCCCAAGTTCGCGGAGGACATCAGTGGAGGCGGCGGCGGCGGCGGCGG 191
51 AlaThrProArgLeuPheProAspAspProArgGluGlyGlnValLysG 67
|||||
192 GCCACACCCCGGCTCTTCCAGATGACCCAGCGGAGGCGGCGGCGGCGG 241
67 nGlyLeuLeuGlyAspCysTrpPheLeuCysAlaCysAlaLeuGlnL 84
|||||
242 GGGGCTGTGGGGGATGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 291
84 ysSerArgHisLeuLeuAspGlnValIleProProGlyGlnProSerTrp 100
|||||
292 AGAGCAGGACCTCTCTGGACAGGTCATCTCTCCGGGACAGCCGAGCTGG 341
101 AlaAspGlnGluTyrArgGlySerPheThrCysArgIleTrpGlnPheG 117
|||||
342 GCCCAGCAGGAGTACCGGGGCTCTTCCCTGCTGCTGCTGCTGCTGCTG 391

117 YArgTrpValGluValThrThrAspAspArgLeuProCysLeuAlaGlyA 134
|||||
392 AGCGCTGGGTGGAGGTGACACAGATGACCGCTGCCGTGCCCTTGACGGA 441
134 rGLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
|||||
442 GACTCTGTTCTCCCGCTGCCAGAGGAGGATGTGTCTGGCTCCCTTA 491
151 LeuGluLysValTyrAlaLysValHisGlySerTyrGluHisLeuTrpAl 167
|||||
492 CTGGAAAAAGGTCTACGCCAAGGTCCATGGTCTCTACGAGCCTCTGGGC 541
167 aGlyGlnValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGluA 184
|||||
542 CGGGCAGGTGGCGATGCCCTGGTGACCTGACCGGGCGGCTGGCAGAAA 591
184 rGTrpAsnLeuLysGlyValAlaGlySerGlyGlyGlnGlnAspArgPro 200
|||||
592 GATGCAACTGAAGGCGCTAGCAGGAAGGAGGCGGCGGCGGCGGCGGCGG 641
201 GlyArgTrpGluHisArgThrCysArgGlnLeuLeuHisLeuLysAspG 217
|||||
642 GCGCGCTGGGAGCACAGGACTTGTGCGCAGCTGCTCCACCTGAAGGACCA 691
217 nCysLeuIleSerCysCysValLeuSerProArgAlaGlyAlaArgGluL 234
|||||
692 GTGCTGATCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 741
234 euGlyGluPheHisAlaPheIleValSerAspLeuArgGluLeuGlnGly 250
|||||
742 TGGGGAGTTCATGCCCTTCATGCTCGGACCTGCGGGAGCTCCAGGGT 791
251 GlnAlaGlyGlnCysIleLeuLeuLeuArgIleGlnAsnProTrpGlyAr 267
|||||
792 CAGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 841
267 gArgCysTrpGlnGlyLeuTrpArgGluGlyGlyGlyGlyGlyGlyGly 284
|||||
842 CGGTGCTGGCAGGGGCTCTGGAGAGGGGGGTGAAGGGTGGAGGCCAGG 891
284 alaPheAlaAlaAlaSerGluLeuLeuSerGlnLeuGlnGlnGlyGlu 300
|||||
892 TAGATGACAGCGGTAGCATCTGAGCTCTCTGCCAGCTCCAGGAGGGGAG 941
301 PheTrpValGluGluGluPheLeuArgGluPheAspGluLeuThrVa 317
|||||
942 TTCGGGTGGAGGAGGAGGAGTTCCTCAGGAGTTTGACGAGCTCACCGT 991
317 lGlyTyrProValThrGluAlaGlyHisLeuGlnSerLeuTyrThrGluA 334
|||||
992 TGGTACCGGCTCAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1041
334 rGLeuLeuCysHisThrArgAlaLeuProGlyAlaTrpValLysGlyGln 350
|||||
1042 GGCTGCTCTGCATACGCGGCGCTGCTGCGGGCTGGGTCAAGGGCCAG 1091
351 SerAlaGlyGlyCysArgAsnAsnSerGlyPheProSerAsnProLysPh 367
|||||
1092 TCAGCAGAGGCTCGCGGACACACAGCGGCTTCCAGCAACCCCAATT 1141
367 eTrpLeuArgValSerGluProSerGluValTyrIleAlaValLeuGlnA 384
|||||
1142 CTGGCTGGGGTCTCAAAACCGAGTGGGTGATATGCTGCTGCTGCTGCTG 1191
384 rGSerArgLeuHisAlaAlaAspTrpAlaGlyArgAlaArgAlaLeuVal 400
|||||
1192 GATCAGGCTGACGCGGCGGACTGGGCGGCGGCGGCGGCGGCGGCGGCGG 1241
401 GlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHisTy 417
|||||
1242 GGTGACAGTATCTCTGTTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1291
417 rGlnAlaValClyLeuHisLeuTrpLysValGluLysArgValAsnL 434

```

1292 CCAGCTGTGGTGTGACCTGTGAAG..... 1319
434 euProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAla 450
1319 ..... 1319
451 TyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTyrTyrLe 467
1319 ..... 1319
467 uAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeuAla 484
1319 ..... 1319
484 rGValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAla 500
1319 ..... 1319
501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyGluTyrGlyThrVa 517
1319 ..... 1319
517 GluLeuArgGlySerThrArgValGlyGluThrAlaGlyGlySerArgA 534
1319 ..... 1319
534 snPheAlaSerTyrProThrAsnProCysPheProPheSerValProGlu 550
1319 ..... 1319
551 GlyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArgPr 567
1319 ..... 1319
567 oSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValProG 584
1320 ..... GTCCGAG 1326
584 IuGlyGlyArgSerGlnAspAlaProProLeuLeuLeuGlnGluProLeu 600
1327 AGGgTGGAGAGCCAGACGACCCGCCACTGCTGTGCTGACAGAGCGGTG 1376
601 LeuSerCysValProHisArgTyrAlaGlnGluValSerArgLeuCysLe 617
1377 CTGAGCTGGCGTCCACATCGCTACGCCAGAGAGGTGAGCCGGCTGTGCT 1426
617 uLeuProAlaGlyThrTyrLysValValProSerThrTyrLeuProAspT 634
1427 CCTCGCTGCAGAGCACTACAGAGTGTGCTCCCTCACCTACCTGCCGAGA 1476
634 hrGluGlyAlaPheThrValThrIleAlaThrArgIleAspArgProSer 650
1477 CAGAGGGGGCTTCACAGTGAACATCGCAACAGAGATTGACAGGCCATCC 1526
651 IleHisSerGlnGluMetLeuGlyGlnPheLeuGlnGluValSerValMe 667
1527 ATTACAGAGCCAGAGATGCTGGGCCAGTCTCTCCAGAGAGTCTCGGTAT 1576
667 tAlaValMetLysThr 672
1577 GGCAGTGTGTAACAC 1592

```

seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-09-422-869-1

seq_documentation_block:

```

; Sequence 1, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: ODA, NAOHISA
; APPLICANT: COX, NANCY J.

```

```

; APPLICANT: SREENAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: OTANI, KENICHI
; APPLICANT: HANIS, CRAIG L.
; APPLICANT: BELL, GRAEME I.
; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES,
; FILE REFERENCE: ARCD:307
; CURRENT APPLICATION NUMBER: US/09/422,869
; CURRENT FILING DATE: 1999-10-21
; EARLIER APPLICATION NUMBER: 60/134,175
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; LENGTH: 49136
; TYPE: DNA
; ORGANISM: Human
US-09-422-869-1

```

```

alignment_scores:
  Quality: 1989.00      Length: 3020
  Ratio: 3.224          Gaps: 11
  Percent Similarity: 20.430  Percent Identity: 20.298

```

alignment_block:

US-09-768-877-2 x US-09-422-869-1 ..

Align seq 1/1 to: US-09-422-869-1 from: 1 to: 49136

```

47 GlnGluIleCysAlaThrProArgLeuPheProAspAspProArgGluI 63
3810 CAGGAGATTGTGTCACACCCCGCTGTTCAGATGACCCAGGAGAGG 3859
63 yGlnValLysGlnGlyLeuLeuGlyAspCysTrpPheLeuCysAlaCysA 80
3860 GCAGGTGAAGCAGAGGGGCTGTGGGGGATGCTGCTGTGTGCTGTGCTG 3909
80 IaAlaLeuGlnLysSerArgHisLeuLeuAsp..... 90
3910 CCGGCGTCCAGAAAGACAGGACCTCTGTGACAGAGTGCAGGAGCCCTTC 3959
90 ..... 90
3960 CCTGTGTTGCTCTGAGCGCGGTTCTTTTGTGCTTCTCCAGCTGTG 4009
90 ..... 90
4010 AGTACAGAGAGGCTTGGAAAGCAGAGCTGTGCCGAGCGGAGTCTCT 4059
90 ..... 90
4060 GCTGTGTGGGGAAGCAGAGAGATTCCAGAGCAGAGCTGAGGACTGC 4109
90 ..... 90
4110 ACTCTGTCCCTCTGCTGTGAGGGGGGCTGCTTGCCCTGCCAGAGGCTC 4159
90 ..... 90
4160 CATCAGGAGGTTGGCCCTGCTGTGTGCTCTGACCCCGGAGTCTCAT 4209
90 ..... 90
4210 GGAGTCAGATCACACGTTTAGAATAAGAGACAATGTCCAGCTCACA 4259
90 ..... 90
4260 GGAAGAGGGGCTGCTGCGAGCCTTGTGCTGAGATCTCTCAGACTAG 4309
90 ..... 90
4310 CTGCTGTGTTTCTTACAGGTTTGAATATAAGTTTGCAAAAAGTTATTA 4359

```

90 90

4360 AACCTGTTCTCTGTGGCTAGACAGATACTCTGGAGGAGAAGCCCTTCTCA 4409

90 90

4410 GGTTTTCTTACCTGGGAGTGTACACCGTTTTATGCTTGGCTGTCTGCTA 4459

90 90

4460 AGTGTGTGATTAAATGACAGCGCGTCAACAGGTGACCTCATTCAGAGT 4509

90 90

4510 TTCACATCATGTCACGCGCCCATGTAAGCGTGTACAGTCACTGCGGCTTT 4559

90 90

4560 CAGACATGGTCTTACCAGCTTTGACTTTTTTTTTTAAACGAGAGTGCTA 4609

90 90

4610 AAATCACTGCCATTGTGTTTCTGCGCCGTAAGTGGCAGAGCCAGGACCGC 4659

90 90

4660 ACCAGTGCGGGTGCCAGCCTGCACTCCCCCGATGCTGGGTGAGATGC 4709

90 90

4710 TTACCCCTGAAGAGAGCCCTGCGGTGGACGCTGTGGTGCAAGCAGCTGGC 4759

90 90

4760 CCAGAGTCGGGGCGCAGGCTCCAGCAGCAGAGGGGCTGCTGTTCCTG 4809

90 90

4810 TGTGACGTGTTGCTTGACCCAGCTCGGTCAAGAACTGGGTCACTCATG 4859

90 90

4860 CCCTTGAATGTCACATTTGTTTGGCTTCAGGTCAGATGCTTTAGTGAG 4909

90 90

4910 GGCAGCAGAGTGTGCCGGGATATGTGGCTCCCTCGGTGTGTGCTCTCA 4959

90 90

4960 GTTTTCCATGAGAGTCTGTATTTCATGTGGTGTATCGAGCCCTGTG 5009

90 90

5010 CAGGCGCCGACATCCAGGTGTCCGTAGAGTTCTCTCGACATCCAGGTG 5059

90 90

5060 TCGCGTAGAGTTCTCTGCGACATCCAGGTGTCCGTAGAGTTCTCTCCGA 5109

90 90

5110 CATCCAGGTGTCCGTAGAGTTCTCTGCGACATCCAGGTGTGGTTAGAG 5159

90 90

5160 TTCTCTGCAGACCGGGTGCCGTGGAGACTCAGCTGTGGCCACACCGC 5209

90 90

5210 GGCCGGGACACTGAGATAGCGCGGGTGGTCTTATATCACGCTCGCCTT 5259

91GlnValIleProProGlyGlnProSerTrp 100
5260 TTGCTTCTCCCTGTGTCATGGCAGGTTCATTCTCCGGGACAGCCGAGCTGG 5309
101 AlaAspGlnGluTyArgGlySerPheThrCysArgIleTrpGlnPheGI 117
5310 GCCGACCGAGGTACCGGGGCTCTTCACCTGTGCGCATTTGGCAGATTGG 5359
117 YArgTrpValGluValThrThrAspAspArgLeuProCysLeuAlaGlyA 134
5360 ACGCTGGGTGGAGTGACACACAGATGACCGCTGCCGTGCCCTGCAGGGA 5409
134 rgLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
5410 GACTCTGTTCTCCGCTGCCAGAGGAGGATGTCTTCGCTCCCTT 5459
151 LeuGluLysValTyAlaLys 157
5460 CTGGAAGGCTACGCCAAGTGCCTGTGCTGGGGGTGAAGGCCCTGGC 5509
157 157
5510 CTGGGCAAGTGGGAGCTGCCACTACCATGGCTGCCCGAGGAGGTCTC 5559
157 157
5560 TGTCTCACTCTGGGCTGCAGAGCCCTTCAGTTCTGAGGGTCTGGCAGCT 5609
157 157
5610 CATTCGTGAGTCAGGCTGACAGGCCAGGTGCACAGATTCTCTTTTGGG 5659
157 157
5660 CCTGTGATGGCCACTCCCTGCTTCCCTTCCCTTGTTCCAAGGCCAG 5709
157 157
5710 CGTGGAGTCGTTCTCCACAGAGAACATGTGCGCTCTCTTATTTTAT 5759
157 157
5760 CGGCCCCAGCAGAAGAAGTGTCTTTATATTTGTTGTGGAATGGTTGGG 5809
157 157
5810 ACAGGCAGACTCATGTCTAGTGTGTGGGAGGAGTCAGGCTACCCGACG 5859
157 157
5860 ATACCAACACTTGTGTATACGGTGTCTTGCTGGCTCAGGGGACCAGGACC 5909
157 157
5910 CTCACATGAGTCAATTAATGAATAGCCTTCCCTTTAGAATGCATTGTG 5959
157 157
5960 TTCTTGCCAAAGCACTGGACTGACAGGCAGCGAGGAAAGCTGGTGAAC 6009
157 157
6010 ATGGGAAGGCTGGCTGGTGACATCAGTGCCAGTGAGCCCTTCCATCCCA 6059
157 157
6060 AGGGCTGTTTTAGGAAAACAGGGTTGGAGCTTGAGACCAAGGATGTG 6109
157 157
6110 GGCATCCATAGTTCACCGCCTCTGCCCTGTCTGCTGTGCCACACCGGA 6159
157 157

6160 TGCCAGAGAGTTTGTGTGTGTGGCAGAGACTGCAGGGCGCTCAGCCTT 6209
 157 157
 6210 GCTGCGAAGTAAGCGCTTTGAAGGTGAGGCTAAGCCTTGACTGTGTGAGG 6259
 157 157
 6260 ATGAGAGAAAGGAGAGAGGGAGTAAGAGGTGGGATTGAGGACAGCGTT 6309
 157 157
 6310 GGAGGATTTGGGGTGTCTACAGACCATGGGAATCAGAGAGGGGGCCATGCT 6359
 158ValHisG1 160
 6360 CAATGCCAGAGGCTCACTCCCATGGTATGTGTCCCTAGGGTCCATGG 6409
 160 ySerTyrGluHisLeuTrpAlaGlyGlnValAlaAspAlaLeuValAspL 177
 6410 GTCTTACAGAGACCTGTGGGGCGGAGGTGCGGATGCCCTGGTGGACC 6459
 177 eutHrGlyGlyLeuAlaGluIuArTrpAsnLeuLysGlyValAlaGlySer 193
 6460 TGACCGGGCGGCTTGGCAGAAAGATGAACTGAAAGGGCGGTAGCAGGAAC 6509
 194 GlyGlyGlnGlnAspArgProGlyArgTrpLeuHisArTrpCysArgG1 210
 6510 GGAGGCCAGAGAGAGAGGCCGCGCTGGAGCAGACAGACTTGTGGCA 6559
 210 nLeuLeuHisLeuLysAspGlnCysLeuLLeSerCysCysValLeuSerP 227
 6560 GCTGCTCCACCTGAAAGCAGTGTCTGATCAGCTGCTGCTGCTCAGGCC 6609
 227 roArGAlaGly..... 230
 6610 CCAGAGAGAGGTGAGGACAGTGGCCACATGGAGGGCTGCAGCCAGCGTG 6659
 230 230
 6660 CCCCCACATGCGCAGGCTCAGGGCACACTGATGTTTTTATGTGACTGGCT 6709
 230 230
 6710 ACAGAGCCCTGTCAAGACTAAGTGGGAAGAATAAGCTGTCTCAAGGG 6759
 230 230
 6760 TGTGTGCTCAGTTTGTGACCTTCCCTGCTGTCTCTCCAGAGGAGC 6809
 230 230
 6810 TGGCCCTTCTCCCTGACCAGTCCCTTCCACTAGTGCAGAGGAGAG 6859
 230 230
 6860 AGGTGACCGAGCTCAAAAGCCACATGCTGTGCCATCCCTGGCCAGCTG 6909
 230 230
 6910 GCACCTGGCAAAATCAAAACCTGTTTTTATTTAGTATAGATACAT 6959
 230 230
 6960 TCGTTAAAAACAGTTTGTCTCCAAAAAATGAAGGGCCAGGTGTGTGG 7009
 230 230
 7010 CTCACACCTGTAATCCAGAACTTTGGAGGCTGAGGTGGAGTATCGCT 7059
 230 230

7060 GGAGCCAGAGATTCAGAGACCCGCTTGGGCAACATGGCAAGATCTCAT 7109
 230 230
 7110 CTCTACAAAAAATGAAGAAAAAAATCATAGATGAACCATGTGAC 7159
 230 230
 7160 TTTTACGTCCGCTCTCAGTTTCCATGACACAGCAGGTTTACGTGTCC 7209
 230 230
 7210 CTTCGCAACCGCTTCTCACACTGCACAGAGCGCTGTCAATGTTCGCC 7259
 230 230
 7260 CCATGAGCGGCTTGGCAGAGCTCTTAATCATTCCTCGATGTGTGATA 7309
 230 230
 7310 TTAAGATCTCTGTGTTTATATATATTAATTAAGCAGTATCAATGAACATCT 7359
 230 230
 7360 TTATCATTTTTTCACTACTAGATTAATTTAAATAATGGTTAAAGAT 7409
 230 230
 7410 ATGATATCACAGTAACTGAACAAAGTTGTATAGCGCTGTGTGGGT 7459
 230 230
 7460 CCCCCATAGACAGCCCTGAGATAGGTTTCAAGACAGCTTGAAGAGTTTC 7509
 230 230
 7510 AGAGAGCTGAGAGGGCGTGTACTCCCGCCAGGCCGCTGGGGTGCCAGG 7559
 230 230
 7560 AAGAGCGCTGGCTCAGAGCCTCCACCTGGAGTGAAGGGCGTGTCTG 7609
 230 230
 7610 GTGGCATTAAGTGGGAGCTGTGGTTTTGGTGCTCCAGGGTGGGT 7659
 230 230
 7660 AGCGTAAGTCTAGCACTTCAGGCTTTAGAGAAAGCCCCCAGCAGAGA 7709
 230 230
 7710 GAGATGCACTGGCACTGACTGAGGTGCATTGAGACCTGTGAGGTGG 7759
 230 230
 7760 CAAGGGCGCGCGGTGGGAGAGACTGAAGTCTTCCAGAGACCTCTCT 7809
 230 230
 7810 CTCTAGACCTTGGCCCGCAGAGATGTGCTTCCCTCTCTGTGTGGCTC 7859
 230 230
 7860 TTTCTTTTATAGCCCATCTGTCTCAAGTCTGTGGCAGAGTGAAGCTGT 7909
 230 230
 7910 TGCCCCCTGGGCTTCAGAGCTTCAGATGGGCTTACAGCAGGCACTAT 7959
 230 230
 7960 CTGATGTCTACTCTAGAGGCTTTGCCAAGAAAGAACTCTAGAT 8009

230 230
8010 TTCTGCACTGTCAGATGATCTGTGCAGCATGTTTGAGAACTGGTGCTG 8059
230 230
8060 TGCCAGGCATGCTGTCACAGATGGGTACGACCATGACTGAGCCCTAGAGA 8109
230 230
8110 AGGGGATGACTTACCAGCACCAGACCCGGATGGCAGCCGAGTCAGGCC 8159
230 230
8160 CGTGTCTTCTCTGGAGATGCTCTGAGCTGATGGGTTCACAGCTGCTA 8209
230 230
8210 AGAAGAGACTCTGCTGTGTCATGGGATCTGGGGTCTCCCTGTAGCCA 8259
230 230
8260 TAGTGGGTGGCTGGAGCATCTCCAGAGAGGAGCAGAGCCCTGTGTG 8309
230 230
8310 TCCTTGTCACTTTGGGACTCCATGTTGCCCTTCTGCTGTGCTCGGCC 8359
231 Ala: ArgGluLeuGlyGluPheHisAlaPheIleVa 242
8360 ATTCTCATGCAAGTGCCTGGGAGCTGGGGAGTTCCATGCTTCATGT 8409
242 lSerAspLeuArgGluLeuGlnGlyGlnAlaGlyGlnCysIleLeuLeuL 259
8410 CTCGGACCTGCGGAGCTCCAGGGTCAGGGGCCAGTGCATCTCTGCTGC 8459
259 euArgIleGlnAsnProTrpGlyArgArgCysTrpGlnGlyLeuTrpArg 275
8460 TCGGGATCCAGAACCCCTGGGGCCGGGTGCTGGCAGGGCTCTGGAGA 8509
276 GluGly..... 277
8510 GAGGGGTGAGTCTGGGGCTGGACCATGCTGCTCGGGAGGGGGGCC 8559
277 277
8560 AGTGCCAGTCTGGCCTGTGCTCCTGCTCACCTTCAGCTGTGACGACTGTAC 8609
277 277
8610 TTGGCTGTCTCAGCAAGGCCCTCAGTCTCCTGCTCTGCTGACACCATGC 8659
277 277
8660 TTGCTTGGCTCCAGGCAATCTTGTGAGGCTGGGACCAGAGTGGCCAT 8709
277 277
8710 TGGGCTGGGGTTTCAATAGGCGACATCATCAGGGCTCGGGCAGCAG 8759
277 277
8760 TCCTGGGAAGAGCATCCAGAGCGGTGAAGTTCTCTGGGAAAAGAGAGG 8809
277 277
8810 GCTCCAGGGTGGCCCTGCCAGAGGCCCTGTGCTGCAGAGCTGCTTCG 8859
277 277
8860 GGTGTGGAGGGGTGCAGAGCTGGGGCAGCGGGTCTGCTGCAGGAACTC 8909

277 277
8910 AGGGCTCTCTGGTCCCTCCAGGCTTCCCCCAGCCTGCCGCAAGTTG 8959
277 277
8960 ACACCTACCACTTCTCGGAGGGCTTCTCTGAGATGAGGTTTCTTCCAG 9009
278 .GlyGluGlyTrpSerGlnValAspAlaAlaValAlaSerGluLeuLeuS 294
9010 GGGTGAAGGTGGAGCCAGGTAGATCAGCGGTAGCATCTGAGCTCTGT 9059
294 erGlnLeuGlnGlyGluPheTrpValGluGluGluGluPheLeuArg 310
9060 CCCAGCTCCAGGAAGGGAGTTCTGGGTGGAGGAGGAGTTCCTCAGG 9109
311 GluPheAspGluLeuThrValGlyTrpProValThrGluAlaGlyHisLe 327
9110 GAGTTTGCAGAGCTCACCGTTGGCTACCGGGTCACGAGGCCGC.CACCT 9158
327 uGlnSerLeuTyThr..... 332
9159 GCAGAGCCTTACACAGGTAGTGCCTCCGAGGGGCTGTGCTGGGCACGTGC 9208
332 332
9209 TCTGCTGCGGAAGTGCAGGAGCTGGGCACGTGCTGGTTCCTCCCTGC 9258
332 332
9259 CCAGGCCAGTTGGTTCTTTCAGCGTGGAGAGATGATTCTGTCCAGG 9308
332 332
9309 ACCCGGGAGGAGGTGATGATTCTGTCCAGGAGCTGGGAGGAGGTGG 9358
332 332
9359 CTGTGTGGAGGGCTGGCTCTGCTGTGCGCTAGTCTGTCTAGACCC 9408
332 332
9409 TCCCAGGGTTTCATGAGGCCACCGTGGCGGAGGCCAGGAGCGGTGT 9458
333 GluArgLeuLeuCysHisT 339
9459 CCCACAGCTGATGCTCTGGTGTCTTCTACTAGAGAGCTGCTCTGCCATA 9508
339 hrArgAlaLeuProGlyAlaTrpValLysGlyGlnSerAlaGlyGlyCys 355
9509 CGCGGGCTGCTGCTGGGGCTGGGTCAAGGGCCAGTCAGCATGAGGCTGC 9558
356 ArgAsnAsnSerGlyPheProSerAsnProLysPheTrpLeuArgValse 372
9559 CGGAACAACAGCGGCTTTCCAGCAACCCCAAAATCTGCTCGGGTCTC 9608
372 rGluProSerGluValTyrIleAlaValLeuGlnArgSerArgLeuHisA 389
9609 AGAACCGAGTGAGGTGTACATTGCCCTCTGCAGAGATCCAGGCTGCAGG 9658
389 laAlaAspTrpAlaGlyArgAlaArgAlaLeuValGlyAspSerHisThr 405
9659 CGCGGACTGGGCGAGCCCGGGCCCGGCACCTGGTGGGTGACAGTCATCT 9708
406 SerTrpSerProAlaSerIleProGlyLysHisTyrGlnAlaValGlyLe 422
9709 TCGTGGAGCCCGAGCATCCCGGGCAAGCACTACCAAGGCTGTGGTCT 9758
422 uHisLeuTrpLysVal..Glu..... 428
9759 GCACCTCTGGAGGTTACTACCCCTGCTGGCTCACGCTCGGTTCACGA 9808
428

9809 GGTGTGTGAGGCCCATGAGGTCTGGGTTCTAGAACTGGCTCTGCCG 9858
428 428
9859 GACACATGTACTCTGCCAGCGGCCCAACAGTCTCCCTCTTGGGC 9908
428 428
9909 TGTGCAAGGGGTGAGCTCTGCTGCTCCAGACCCGGCTGTACT 9958
428 428
9959 GAGACTCAGGTCTTCTGCCTTGCCTGTGCTCATTTGGCCAAAGAAA 10008
428 428
10009 GCAACAGATCTGCAGCGAGGGCAGAACCCGAGAGGGGCTTGACCCG 10058
428 428
10059 GGGGGCTCTGGACGCCCTGCTTCTAGGCAAGAGGTAGAGCCAG 10108
428 428
10109 CGGCTGAGACCTGTCAAGGCCAGTCCAGCTCTGCAGCTTGCTGTGTA 10158
428 428
10159 CTTGGACACATCTCTCTCCCTGCTCCAGTCTTCTCCCTGCAAGAC 10208
428 428
10209 GGGGTCTGCACAGGATCTATGGGATGCTGTGAGGCCAGGACATCCC 10258
428 428
10259 AGGCTCAACCATCTGTTCACAAAGTGTGTTTCCAGAAACAGATG 10308
428 428
10309 GGGGCCCTGAGGGCAAGGGCTGAGTGTGTCAGAGATATGCCGGCTG 10358
428 428
10359 CTGCTCAGGGGCTGGGTTTTCATCTTGTCTGTGCTTGACAGGGTGTACA 10408
428 428
10409 CTTGGACACACTGTTCCCTGCTCATGATGTGGCCACATGATG 10458
428 428
10459 TTCCTTCTCTTGCAAAAGAAAGTCTGGAAGGCCACTGTCCACAGC 10508
428 428
10509 CCCAGGTTGGCTGGGCCACAGTGCCTTGTGGGCCACGCTACAGAGAG 10558
428 428
10559 ACTTGAGGCTGCTGTCTGGGACAGATACGTGGCCAGGCAAGTGAAG 10608
428 428
10609 CCCGGATTGGTGGCATCTTAGCTGTCTGAGAGAGGGTGAAGGT 10658
428 428
10659 GGTGACAGGCCCTTGGGCTTTTCATCTGTCACTCCAGAGGCCCTTGTCT 10708
428 428

10709 TGCAGCAGGAGGTCAAGGCCAGGGGCTGTGACCCCGCCCTCTCCAC 10758
429 LysArgValAsnLeu 434
10759 ACTGAGCTCTGCACGTCTCACAGGTAGAGAAAGCGGGTCAATCTG 10808
435 ProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAlaTy 451
10809 CTTAGGGTCTCTCCATGCCCCCGTGGCTGGCACCGTGTCCATATGATA 10858
451 RAspArgGluValHisLeuArgCysGluLeuSerProGlyTyrTyrLeu 468
10859 CGACCGGAGGTCCACTGCGCTTGTAGCTCTCACCGGGCTACTACTGG 10908
468 LaValProSerThrPheLeuLysAspAlaProGlyGluPheLeuArg 484
10909 CTGTCCCGACACTCTCTGAAGACGGCCAGGGGAGTTCTCTCTCGA 10958
485 ValPheSerThrGlyArgValSerLeu..... 493
10959 GTCTTCTTACCGGGCGAGTCTCCCTTAGTGTAGAGAGAACCGCGACTGC 11008
493 493
11009 TGTGCTCTCCAGGCCACAGGCCCTTCCAGCAGAGATTGGCACTT 11058
493 493
11059 TCCTCTGTGGTGGCAGGTGTCCATGTGGAACTGAGGCCACCTGGAC 11108
494 SerAlaIleArgAl 498
11109 CTGCTGCCAGCCCTCCCATGTTTGTCTTCTTGGCAGCGCATCAGGGC 11158
498 aValAlaLysAsnThrThrProGlyAlaAlaLeuProAlaGlyLurPG 515
11159 AGTGGCCAAAGACACACCCCGGGGCAAGCTCTCTCGGGGAGTGGG 11208
515 LThrValGlnLeuArgGlySerTrpArgValGlyGlnThrAlaGly 531
11209 GGAACGTGACGTACGGGGTCTTGAGAGTGGCCAGAGCGGGGGGC 11258
532 SerArgAsnPheAlaSerTyrProThrAsnProCysPheProPheSerVa 548
11259 AGCAGGAACCTTGGCTCATACCCCAACCCCGCTTCCCTTCCTGCTG 11308
548 LProGluGlyProGlyProArgCysValArgIleThrLeuHisGlnHis 565
11309 CCCGAGAGGGCTGTGGCCCGCTGCGCTGCGCATCTCTGATCAGACT 11358
565 ysaArgProSerAspThrGluPheHisProIleGlyPheHisIlePhe... 580
11359 GCCGGCCACTGACACCGAGTTCACCCCATTCGGCTTCATCTTCCAG 11408
580 580
11409 GCAAGCTCTTGGCCAGGAGGAGGAGGAGAGAGAGGGCCCTCAGAG 11458
580 580
11459 AATTGCACTTGGCCCTCCATTGTCCACAGAGGGCTGTGGCTCAGTC 11508
580 580
11509 ACTTGGGCTCCCTGCTTGCAGGGCTGCTAGAACCCGACAGGGC 11558
580 580
11559 CTTCTCCATCTCCAACTCTCAGAGGCAAGGCCAAGATGGCTGTGGA 11608
580 580
11609 AGGGCGGGGGCTTGGAGGTGGGCAAGGCTGATCCAGGAGGCAAGTT 11658

580	580	580
11659	TCACAGGAGGTGTTGACTGGGGAGGAAGGAGAGTATTGGAGAGACAG	11708		
580	580	580
11709	GAGGCGAGGTTGAGACCAGCGGGGTGGTTCGAGCCCTGCCTTGGGAAC	11758		
580	580	580
11759	GCAGGGGCTGATGACTCAGGAGTGCAGAGGGGAGGCCACAGCTGGC	11808		
580	580	580
11809	TGGCCACACGACCCCTCGGTGTGAGGAAGTCCACAGTCACTGAGCTCA	11858		
580	580	580
11859	GCCAGCAGCCCTGTCCACTTACCTGACTCAGAATGACTGTGTCCCAAG	11908		
580	580	580
11909	GPTCAATCTCTGCACACATGTCTCCCTGGAAATCAGGGGCCCTGACGAG	11958		
580	580	580
11959	GAAGCACTGCAACCCCTCGGTTACAGTGGGCTGCCTGGGACCCCTTGA	12008		
580	580	580
12009	CCCTCGCTGTTTGCCCTGGGCCACCGCTCAGGTCCCTTAGAGTCTGAG	12058		
580	580	580
12059	GAACAACATGCCAGGGCCAGTGGGAGCCCTTGGGGCGGGCTGGGCAGTC	12108		
580	580	580
12109	ACAGTGTTAAAGCCCTGATGATGTGACAGGCCCTCCAGGGGGGGCCCC	12158		
580	580	580
12159	ACTGCCGCACCTTCTGCAAGGCTGGCAGGCCCTTGGTAGGAGGCGAG	12208		
580	580	580
12209	TCCAGTGTCCAGGCTGCGACGCCCTCTCAGAGAGGGGCTGTATGTGA	12258		
580	580	580
12259	CTCAAGAGGGCCAGGGGATCCGAGCAGATGGCCCTGGGCTGGGCTCCCT	12308		
581	581	583
12309	ACCCCAAGGCTGGCCCCCTCAGTCTGAGCCTGCGCTTTCTCCTCAGGTCCC	12358		
583	oeluclyglyArgSerGlnAspAlaProProLeuLeuLeuGlnGluProL	600		
12359	AGAGGTTGGAGGAGCCAGGACGACCCCACTGCTGCTCAGAGAGCGCC	12408		
600	euLeuSerCysValProHisArgTyrAlaGlnGluValSerArgLeuCys	616		
12409	TCCTGAGCTGGTGCCACATCGCTACGCCCCAGGAGTGCACCGGCTCTGC	12458		
617	LeuLeuProAlaGlyThrTyrLysValValProSerThrTyrLeuProAs	633		
12459	CTCTGCTCGGGCACCCTACAAGTTGTGCCCTCCACCTACCTGCCGA	12508		
633	pThcIuGlyAlaPheThrValThrIleAlaThrArgIleAsp	647		
12509	CACAGAGGGGCTTCACAGTGACCATCGCAACACCGAGATTGACAGTTGG	12558		

```

647 ..... 647
12559 GCTCTGGGACTTGGGGGGCCAGCTGGAGGCTGGGGTCTGGAGTCTTA 12608
647 ..... 647
12609 GFGCTGGCGTGTCCCCCACAGTCTCCTGCCCTGCGCCTCACCTCAAGCCC 12658
647 ..... 647
12659 CTATCTGCTTGGCAGACCGGCTGTCTGCCTACCTGGGGACCCCTTCC 12708
647 ..... 647
12709 TTGCTGGTCTGAGCCTGGAAGGAGAGTCTAGTGGGAGGTGGCCAGGAGC 12758
647 ..... 647
12759 ACACAGCCACTGTGTGACAAGTCAGTCTGGGAGCGTGATCTGGTGTC 12808
648 .....ArgProSerIleHisSerGlnGluMetIleuGlyGlnPheLeuG 662
|||||
12809 TCTCCACAGGCGCATCTCATTCACGCCAGGAGATCTGGCCAGTTCCTCC 12858
662 InGluVal 664
|||||
12859 AAGAGGTG 12866

seq_name: /cgn2_6/ptodata/2/ina/5B_comb.seq:US-08-835-099A_5

seq_documentation_block:
; Sequence 5, Application US/08835099A
; Patent No. 5874277
; GENERAL INFORMATION:
; APPLICANT: SHINTANI, Yasushi
; APPLICANT: NISHI, Kazuori
; APPLICANT: KAWAMOTO, Tomohiro
; TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION
; TITLE OF INVENTION: AND USE
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/835,099A
; FILING DATE: 04-APR-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 083649/1996
; FILING DATE: 05-APR-1996
; APPLICATION NUMBER: 97105508.2
; FILING DATE: 03-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Resnick, David S
; REGISTRATION NUMBER: 34,235
; REFERENCE/DOCKET NUMBER: 47342
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6240
; TELEX:
; INFORMATION FOR SEQ.ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2109 base pairs
; TYPE: nucleic acid

```


STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 US-08-835-099A-5

Alignment_scores:
 Quality: 635.00 Length: 508
 Ratio: 2.096 Gaps: 17
 Percent Similarity: 59.646 Percent Identity: 32.283

alignment_block:
 US-09-768-877-2 x US-08-835-099A-5

Align seg 1/1 to: US-08-835-099A-5 from: 1 to: 2109

```

13 Leupheargaspalaalaphroalaalaspserleu...phecy 28
|||||
133 CTATTAGAGACCTGAGTCCACATGTCATGAGCTTGCGCTACAA 182
28 sasleuserthrproleualaglnphearglupspilthrparga 45
|||||
183 GGATCTTGA...CCAGCTCTCCCAACATCAAGCATCTGGAAGC 229
45 rproglngluilecysalathrproargleupheproaspasproarg 61
|||||
230 GGGCCAGGAGTTGTGTCAGCCCTCAGTTATCTGTGTGAGCCACG 279
62 gluglglnglvalylsngllyleuleugllyaspysrtrpbeleucysal 78
|||
280 CGCACAGACATTTGTCAGGAGTGTCTAGCTGACCTGCTTGCTGCTGC 329
78 acysalalaleuglnlyseratghisaleuleuaspalnval11leprop 95
|||||
330 CATTCCTCTCCAGCCCTGAAATGAAGCTCTTACCGGGTGGCCCCA 379
95 roglnglnprosertrpalaaspnglnlytrarglyserphethrcys 111
|||
380 GGGACACAGAC...TTCCAGAGACATGCGGGAATCTTCTACTTT 423
112 Arg11letrpnlphelgylargtrpvalglvalalthrthrasparagle 128
|||||
424 CAGTTCTGGCAGTAGCAGAGTGGTGGAGCTTGCATTGACGACAGCT 473
128 uprocysleualaglyargleucyspheserargcysglalnarglupspv 145
|||||
474 GCCCACCAGAAATGACAGCTGCTCTCTACCTCGAACACAAAGCAATG 523
145 alpheatrleuproleuleuglnlyvaltrpvalylsvalhisglyser 161
|||||
524 AATTCTGAGTGCCTCTGAGAAAGCTTATGCAAGCTTAAAGGTTGT 573
162 tyrglnhisleutrpalaaglglvalalaspalaaleuvalasprleuth 178
|||||
574 TATGAGGCTCTCGCTGAGGTTCCACAGCTGAGAGGCTTGAAGATTTCAC 623
178 rglglglnglvalglargtrpansleuylsvalalaglyserglg 195
|||||
624 AGGTGGATCTCTGAGTTTATGACCTGAG... 654
195 lylnglnlaspargprogllyargtrpnlhisargthrcysarglnleu 211
|||||
655 .....AAACCAACAGCAATCTATATCATGATCATCCGAAAGGCC 693
212 leuhalileuylsaspnglncys.....leu11esercyscysal.. 224
|||
694 CTC.....TGTGGGGGTCTCTGCTGGGCTCTCCATTCGA 728
225 leuserproalaglalalyala.....Arggluleugllyg 236
|||||
729 TGTCTCCAGTCCAGCCGAGCCGAAAGCCATCACCAGCAGCAAGACTGTTTA 778
236 lurbenlialaphelilevalseraspaleuarggluleuglnglvalala 252

```

```

779 AGAGTCATGCTGACTGTCTGAGTGCAGAGTGAAGTGAATTTCCAG... 825
253 glylncyslileuleuleuarg11eglnasprotrpalyargargcy 269
|||||
826 GCCATCCAGAGAGCTGATCAGACTCAGAAATCCATGAGGCTGAAGTGA 875
269 strpngllyleutrpargglugllygllygllytrpserglvalaspra 286
|||||
876 GTGTGCGGAGCTTGAGCATGATGACACAGAGTGAATCAATCAATGAGCC 925
286 lalavalalasergluleuleuserglnglglnglgluphetp 302
|||||
926 CCCGGCGAAGAGAGACTGACAGCAAGAAAGTTGAGATGAGAAATTTCTG 975
303 valglnglglnglupheleuarglupheaspgluleuthrvalgllyty 319
|||||
976 ATGTACTTTCAGATTCTGAGCAGATCTCTCGGTGAGATCTGCA 1025
319 rprovalthrualaglhisleuglnserleuyltrhglnuargleu 336
|||||
1026 CCTGTCCCGGAGCTCTGAGTAGCAGAGAGGTGCAAAATGAACTTG 1075
336 eucysnlsrthrargalaleuoprogllyalatrvalylsnglinsera 352
|||||
1076 TCCTGTTCAC.....GGCCACTGACCCCGGGCTCCACAGCT 1113
353 glyllycysargasnserglyphepro.....Serasnpr 365
|||||
1114 GGGGGCTCCAGAAC.....TACCAGCAGACTGACGACCAATCC 1154
365 olyspheutrleuargvalsergluproserglu..... 376
|||||
1155 CCAGTCAAAATCCCTTGGATGAAGTGAAGAGACAGAGAGAGACA 1204
377 .....Valtyrilealavalaleuglnaraserarg 386
1205 TCGGTGAACCTGCTGACAGTGTCTGGGCTGTGATGCAAGAAATGCG 1254
387 leuhalalalasprrpalaagllyargalalargalaleuvalglaspse 403
|||||
1255 .....AGGTGGCGAAGCCGAAATGAGCAAGG 1280
403 rhlsrsertrpserproalaserleprogllylshisyrnglnalav 420
|||||
1281 CATGCTTAC.....A 1291
420 alglyleuhalileutrpalyvalglupysarg.....Arg 431
|||||
1292 TCGGCTATGCGCTGTACAGGTTCCAAAGAGCTGAGAGTACACAGAC 1341
432 valasnlleuproarg.....Valleusermetprovalalaglyth 446
|||||
1342 GCACACTTGGCCCGGATTTCTCTCTGCTTACAGCCCTCAGCCGCGAC 1391
446 lalacysnlsalatrargparggluvalhisleuargcysgluleuser 463
|||||
1392 CAGACCTTACGTAACCTGCGGAGGTCTGCGCGGCGCGGCTGCGCC 1441
463 rogllytyrtyrleualavalproserthrphelileuylsaspala 479
|||||
1442 CTGGGAGTAGTACCTGTGTGCTGATCATTGAAACCTTCAAGAGAGCC 1491
480 glupheleuargvalpheser 487
|||||
1492 GAGTTCTGTGAGAGTGTCTCA 1515
seq_name: /cgn2_6/ptodata/2/lna/6A.COMB.seq:US-09-157-349-5
seq_documentation_block:
; Sequence 5, Application US/09157349
; Patent No. 6068990
; GENERAL INFORMATION:

```

APPLICANT: SHINTANI, Yasushi
 APPLICANT: NISHI, Kazuori
 APPLICANT: KAWAMOTO, Tomohiro
 TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION
 TITLE OF INVENTION: AND USE
 NUMBER OF SEQUENCES: 18
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
 STREET: 130 Water Street
 CITY: Boston
 STATE: MA
 COUNTRY: USA
 ZIP: 02109
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/157,349
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/835,099
 FILING DATE:
 APPLICATION NUMBER: 97105508.2
 FILING DATE: 03-APR-1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Resnick, David S
 REGISTRATION NUMBER: 34,235
 REFERENCE/DOCKET NUMBER: 47342
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617-523-3400
 TELEFAX: 617-523-6440
 TELEX:
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2109 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 US-09-157-349-5

alignment_scores:		
Quality:	635.00	Length: 508
Ratio:	2.096	Gaps: 17
Percent Similarity:	59.646	Percent Identity: 32.283

alignment_block:

US-09-768-877-2 x US-09-157-349-5

Align seg 1/1 to: US-09-157-349-5 from: 1 to: 2109

[illegible]

330 CATTGGCTCCCTGACCGTGAATGAAGAGCTGCTTTAACC GGGTGGTCGCCCA 379

95 roGlyGlnProSerTrpAlaaspGlnGluThrArgGlySerPheThrCys 111
||| : : : : : ||| ||| |||

380 GGGACCAAGGAC TTCCAGAGAACAATATCGGGAATCTTTCACTTT 423

112 ArgIleTrpGlnPheGlyArgTrpValcIuVaiThrThrAspAspArgLe 128
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||

424 CAGTCTCGGCAGTACGAGAGTGCGTGGAGGTGGTCTATTGACGACAGGCT 473

128 uProCysLeuAlaGlyArgLeuCysPheSerArgCysGlnArgGluaspv 145
||| ||| : : : : : ||| ||| : : : : : ||| : : : : : ||| : : : : : |||

474 GCCCACCAAGAATGCAGAGCTGCTTCTCATACACTCGGAACAAGCAATG 523

145 alpheTrpLeuProLeuGluLulysValThrAlalysvalHisGlySer 161
||||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||

524 AAITCTGGAGTGCCCTGCTCGAAGAACCCATATCCAAAGCTTAATGGTTGT 573

162 TyrGluHisLeuTrpAlaGlyGlnValAlaaspAlaLeuValaspLeuth 178
||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||

574 TATGAGCTCTCGTGGAGTTCACAGTGGAGGGTTTGAGGATTCAC 623

178 rGlyGlyLeuAlaGluArgTrpAsnLeuLysGlyValAlaGlyserGlyg 195
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||

624 AGTGGCATCTCGAGTTTATGACCTGAAG 654

195 LyGlnInAspArgProGlyArgTrpGluHisargThrCysArgGlnLeu 211
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||

655 AAACCACGAGCCAATCATATCAGATCATCCGGAAGGCC 693

212 LeuHisLeuLysaspGlnCys LeuIleSerCysCysVal . . . 224
||| ||| : : : : : ||| : : : : : ||| : : : : : |||

694 CTC TGTCGGGGTCTCTGCTGGGCTGCCATTGA 728

225 .LeuSerProArgAlaGlyAla ArgGluLeuGlyG 236
: : : : : ||| ||| : : : : : ||| : : : : : |||

729 TGTCTCCAGTCGACGCCGAAGCCAGCATCACAGCCAGAAGCTGGTTA 778

236 luPheHisAlaPheIleValSerAspleuArgGluLeuGlnGlyGlnAla 252
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||

779 AGATCATCGGTACTCTGTCACCTGGAGTCCGAAGAGTGAATTTCCAG . . . 825

253 GlyGlnCysIleLeuLeuLeuArgIleGlnasnProtpGlyArgArgCy 269
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||

826 GGCCATCCAGAGAAGCTGATCAGCTCAGGAATCCATGGGGTGAAGTGA 875

269 sTrpGlnGlyLeuTrpArgGluGlyGlyGluGlyTyTrpSerGlnValaspa 286
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||

876 GTGTCGGGAGCCCTGGAGCGATGATGCACCAGATGGAATCATCATAGACC 925

286 laalavalAlaserGluLeuSerGlnLeuGlnGluGlyGluPheTrp 302
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||

926 CCCGGGGAAGGAAGAACTGGACAAGAAAGTTGAGGATGGAGAATTTCTGG 975

303 ValcIuGlugluGluPheLeuArgGluPheAsp6LeuThrValGlyTy 319
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||

976 ATGTCATTTTCAGATTTCTGGAGGCGAGTCTCTCGGTTGGAGATTCGCAA 1025

319 rProValThrGluAlaGlyHisLeuGlnSerLeuTyThrGluArgjeul 336
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||

1026 CCTGTCCCCGGGACTCTGTAGTAGCAGGAGGTGCACAAATGGAACCTGG 1075

336 euCysHistrArgAlaLeuProGlyAlatrPrpVallysGlyGlnSerAla 352
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||

1076 TCCTGTTCAC GGCCATCGACCCGGGGCTCCACAGCT 1113

353 GlyGlyCysArgAsnAsnSerGlyPhePro SerAsnPr 365
||| ||| : : : : : ||| : : : : : ||| : : : : : |||

1114 GGGGGCTGCCAGRAC TACCAGCCACGACTGTGGACCAATCC 1154

365 oLysPheTrpLeuArgValIserGluProSerGlu 376
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||

1155 CCAGTTTCAAAAATCCGTTTGGATCACTGGATGATGAGACCAGGAGGAGCA 1204

```

TELEX:
: INFORMATION FOR SEQ ID NO: 6:
:-----
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 2136 base pairs
:   TYPE: nucleic acid
:   STRANDEDNESS: double
:   TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: US-08-835-099A-6
:-----

alignment_scores:
      Quality: 635.00      Length: 508
      Ratio: 2.096      Gaps: 17
Percent Similarity: 59.646      Percent Identity: 32.283

alignment_block:
US-09-768-877-2 x US-08-835-099A-6 ..

Align seg 1/1 to: US-08-835-099A-6 from: 1 to: 2136

13 LeuPhearGaspaAlaAlaIaphProAlaAlaAspSerSerLeu...PheCy 28
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
160 CTATTTAAGAGCAGCCCTGAGTTCGCCAGCATGTCATCAGCTTTGGCTACAA 209
28 sAspLeuSerThrProLeuAlaGlnPheArgGluAspIleThrTyrArg 45
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
210 GGATCTTGGGA...CCAGGCTCTCCGCAAACTCAAGGCATCATCTGGAAAGC 256
45 rGpProGlnGluIleCysAlaIleThrProArgLeuPheProAspArg 61
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
257 GGCCACAGGAGATTGTGCCAGCCCTCAAGTTATCTGTTGGTGAAGCCAGC 306
62 GluGluGlnValIleCysGlnGlyLeuGluGlyAspCysTyrPheLeuCysAl 78
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
307 CGCAGACAGCATTTGTGAGGGGTGTCTAGGTGAGTCTGCTGCTTGGCTGTC 356
78 acyAlaAlaAlaLeuGlnIleCysSerArgHisLeuLeuAspGlnValIlePro 95
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
357 CATTCGCTCCCTGACCCCTGATGAAGAGCTGTTTACCGGGTGTGCCCA 406
95 roGlyGlnProSerThrPalAspGlnIleLurArgGlySerPheThrCys 111
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
407 GGGACACAGGAC.....TTCAGAGACACTATGCGGAAATCTTCACATT 450
112 ArgIleTyrGlnPheGlyArgTyrValGluValIleThrThrAspAspArg 128
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
451 CAGTTTCGGCAGTACGAGAGAGTGGGTGAGGTGCTATTACGACAGAGCT 500
128 uProCysLeuAlaGlyArgLeuCysPheSerArgCysGlnArgGluAspY 145
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
501 GCCCACCAAGAAATGAGACAGCTGCTCTTCCACCTCGGAGCAAGCAAT 550
145 alPheTyrLeuProLeuLeuGluLysValIleTyrAlaLysValHisGlySer 161
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
551 AATTCTGGAGTGGCCCTGCTGGAGAAAGCCATATCCAAAGCTTAATGGTTGT 600
162 TyrGlnHisLeuTyrPheArgIleValAlaAlaAspAlaLeuValAspLeuThr 178
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
601 TATGAGCTCTCGCTCGAGAGTTCACACAGTCGAGAGGGGTTTCAGATTTCAC 650
178 rGlyGlyLeuAlaGluArgTyrAsnLeuLysGlyValAlaGlySerGly 195
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
651 AGGGGCGATCTCTGAGTTTATGACCTGAAG..... 681
195 LysGlnGlnAspArgProGlyArgTyrPheGlnHisArgThrCysArgGlnLeu 211
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
682 .....AAACCAACCAAGCCAAATCTATATATGATATCATCGGAAGGCC 720
212 LeuHisLeuLysAspGlnCys.....LeuIleSerCysGlyVal... 224
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
721 CTC.....TGTGGGGGGGTCTCTGCTGGGGGTGCTGCATCTTA 755

```

```

225 .LeuSerProArgAlaGlyAla.....ArgGluLeuGlyC 236
    :|||  |||  |||
756 TGCTCCAGTGCAGCCGACCAAGCCATCACCAGCAGCAAGCTGGTTA 805
    :
236 luPheHisAlaPheIleValSerAspLeuArgGluLeuGlnGlyGlnAla 252
    :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
806 AGAGTCAATGCTACTCTGCTCAGCTGGAGTCAAGAGGTGAATTTCCAG... 852
    :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
253 GlyGlnCysIleLeuLeuLeuArgIleGlnAsnProTyrGlyArgArgCys 269
    :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
853 GCCCATCCAGAGAAGCTGATCAGACTCAGCAATCCATGGGTGAAGTGA 902
    :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
269 sTrpGlnGlyLeuTrpArgGluGlyGlyGluGlyTrpSerGlnValAspA 286
    :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
903 GTGGTGGGAGCTGGAGCGATGATCCAGCAGAGTGGGAATCAGATAGACC 952
    :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
286 laAlaValAlaSerGluLeuLeuSerGlnLeuGlnGluGlyGluPheTrp 302
    :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
953 CCGCGCGGAGGAGAGACTGGACAGAAAGTTGAGGATGGAGATTTCTGG 1002
    :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
303 ValGluGluGluGluPheLeuArgGluPheAspGluLeuThrValGlyTy 319
    :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1003 ATGTCACATTTCAGATTCTGTCGAGGAGTCTCTCGGTGGAGATCTGCAA 1052
    :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
319 rProValThrGluAlaGlyHisLeuGlnSerLeuTyrThrGluArgLeuL 336
    :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1053 CTTGTCCTCCGGACTCTCTCAGTAGCAGGAGGTGCACAAATGCAACCTGG 1102
    :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
336 euCysHisThrArgAlaLeuProGlyAlaTrpValLysGlyGlnSerAla 352
    :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1103 TCCTGTTCAAC.....GGCAGCTGGACCGGGCTCCACAGT 1140
    :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
353 GlyGlyCysArgAsnAsnSerGlyPhePro.....SerAsnPr 365
    :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1141 GGGGCTGCCAGAAC.....TACCCAGCCACGACTGTCGACCAATCC 1181
    :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
365 OlyPheTrpLeuArgValSerGluProSerGlu..... 376
    :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1182 CCAGTTCAAATCCGTTTGATGAAGTGGATGAGGACGAGGAGGAGCA 1231
    :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
377 .....ValTyrIleAlaValLeuLeuGlnArgSerArg 386
    :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1232 TCGGTGAACCTCTCTACAGTCTGCTGGCTGATGAGCAAAATCCG 1281
    :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
387 LeuHisAlaAlaAspTrpAlaGlyArgAlaArgAlaLeuValGlyAspSe 403
    :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1282 .....AGGTGGCGGAGCGGATAGGACAAGG 1307
    :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
403 rHisThrSerTrpSerProAlaSerIleProGlyLysHisTyrGlnAlav 420
    :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1308 CATGCTTAC.....A 1318
    :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
420 alGlyLeuHisLeuTrpLysValGluLysArg.....Arg 431
    :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1319 TCGGCTATCGCTCTACAGGTTCACAGGAGTCCAGGAGTCCACAGGAC 1368
    :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
432 ValAsnLeuProArg.....ValLeuSerMetProProValAlaGlyTh 446
    :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1369 GCACACTTGGCGGGGATTTCTTCTGGCTTACAGCCCTCCAGCCGCGAC 1418
    :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
446 rAlaCysHisAlaTyrAspArgGluValHisLeuArgCysGluLeuSerP 463
    :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1419 CAGCACCCTACGTACACCTCGGGAGGTCTCTGGCGGCGCGGCTGCCCC 1468
    :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
463 roGlyTyrTrpLeuAlaValProSerThrPheLeuLysAspAlaProGly 479
    :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1469 CTGGGAGTACCTGGTGGTGCATCCACATTTGAACCTTCAACAGCGC 1518
    :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
480 GluPheLeuLeuArgValPheSer 487
    :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1519 GAGTTCTGCTTGAGATGTTCTCA 1542
    :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq.us-09-157-349-6

```

seq_documentation_block:
; Sequence 6, Application US/09157349
; Patent No. 6068990
; GENERAL INFORMATION:
; APPLICANT: SHINTANI, Yasushi
; APPLICANT: NISHI, Kazunori
; APPLICANT: KAWAMOTO, Tomohiro
; TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION
; TITLE OF INVENTION: AND USE
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09157.349
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/835,099
; FILING DATE:
; APPLICATION NUMBER: 97105508.2
; FILING DATE: 03-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Resnick, David S
; REGISTRATION NUMBER: 34,235
; REFERENCE/DOCKET NUMBER: 47342
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; TELEX:
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2136 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-09-157-349-6

alignment_scores:
Quality: 635.00 Length: 508
Ratio: 2.096 Gaps: 17
Percent Similarity: 59.646 Percent Identity: 32.283

alignment_block:
US-09-768-877-2 x US-09-157-349-6 ..
Align seg 1/1 to: US-09-157-349-6 from: 1 to: 2136

13 LeuPheArgAspAlaPheProAlaAlaAspSerLeu...PheCy 28
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
160 CTATTAAAGGACCTGAGTTCACAGATGTCATCAGCTTTGGGTACAA 209
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
28 sasPleuSerThrProLeuAlaGlnPheArgGluAspIleThrTrpArgA 45
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
210 GGATCTTGA...CCAGGCTCTCCCAACCAAGGCATCATCTGGAAGC 256
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
45 rqProGlnGluIleCysAlaThrProArgLeuPheProAspProArg 61
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
257 GCCCACCAGGAGTGTGTCACCGCTCTCAGTTTATCGTTGGTGGAGCCAGC 306
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
62 GluGlyGlnValLysGlnGlyLeuLeuGlyAspCysTrpPheLeuCysAl 78
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

```

1141 GGGGGGGCGCCAGAAC.....TACCCAGCCAGCTACTGGACCAATCC 1181
365 oluSpHeTrpLeuAlaValSerGIuProSerGIu..... 376
      ::::|||||:|||||:|||||:|||||:|||||
1182 CCAGTTCACAAATCCGTTGGATGATGAGTAAAGGAGACAGCAGAGACGA 1231
377 .....ValTrpIleuAlaValIleuGIuArgSerGIu 386
      :|||:.....:|||||:|||||:|||||
1232 TCGGTGAACCCCTGCTACAGTGGCTGGGCGCTGATGCACAAAAATGCG 1281
387 LeuAlaIleuAlaAspTrpAlaGIuArgAlaValLeuValGIuAspSe 403
      :|||:|||||:|||||:|||||:|||||
1282 .....AGTGGCGGGAAGCGGATAGCAACAAAG 1307
403 rHISrHSerTrpSerProAlaSerIleProGIuLysHISrGlnAlaVal 420
      :|||
1308 CAGCTGTACG.....A 1318
420 alGIuLeuHISLeuTrpLysValGIuLysArg.....Arg 431
      :|||||:.....:|||||:|||||
1319 TCGGGATATGCGGTCGTACACAGGTTCCCAAGAGCTGAGAGTCAACGGCAC 1368
432 ValAsnLeuProArg.....ValIleSerMetProProValAlaGIuTh 446
      :|||||:|||||:|||||:|||||
1369 GCACACTTGGGGCGGGGATTTCTTCCTGGCTACCAAGCCCTTACGCCCGAC 1418
446 rAlaCysHISAlaTrpAspArgGIuAlaHISLeuArgCysGIuLeuSerP 463
      :|||:.....:|||||:|||||:|||||
1419 CAGCACCCTAGCTCAACCTGGGGAGAGCTGTGGCGGGGCCCGCTGCCCC 1468
463 rGGLyTrpTrpLeuAlaValProSerThrPheLeuLysAspAlaProGLy 479
      :|||||:|||||:|||||:|||||:|||||
1469 CTGGGGAGTACCTGGTGGTGCACATTCACAAATTCGAACCTTCAAAGAGCGC 1518
480 GIuPheLeuLeuArgValPheSer 487
      :|||||:|||||:|||||:|||||
1519 GAGTCTCGCTTACGAGATGTTTCA 1542

```

```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-422-869-15
seq documentation block:
? Sequence 15, Application US/09422869
? Patent No. 6235481
GENERAL INFORMATION:
? APPLICANT: POLONSKY, KENNETH S.
? APPLICANT: HORIKAWA, YUKIO
? APPLICANT: ODA, NAOHISA
? APPLICANT: COX, NANCY J.
? APPLICANT: SREENAN, SEANUS
? APPLICANT: ZHOU, YUN-PING
? APPLICANT: OTANI, KENICHI
? APPLICANT: HANIS, CRAIG L.
? APPLICANT: BELL, GRAEME I.
TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
FILE REFERENCE: ARCD:307
CURRENT APPLICATION NUMBER: US/09/422,869
CURRENT FILING DATE: 1999-10-21
EARLIER APPLICATION NUMBER: 60/134,175
EARLIER FILING DATE: 1999-05-13
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 15
LENGTH: 1267
? TYPE: DNA
? ORGANISM: Human
US-09-422-869-15

```

```

alignment_block:
US-09-768-877-2 x US-09-422-869-15
Align seg 1/1 to: US-09-422-869-15 from: 1 to: 1267

1 MetArgAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
|||||
141 ATGCGGGCGGGCGGGCGGCGAGCCGCGGAGGAGCTGTTCGGGACGC 190
|||||
17 aAlaPheProAlaAlaAspSerLeuPheCysAspLeuSerThrProL 34
|||||
191 CGCCTTCCCGCGCGGACTCTCTGCTCTCTCGGACTGTCTACGCCGC 240
|||||
34 euAlaGlnPheArgGluAspIleThrTrpArgArgProGlnGluLeu 50
|||||
241 TGGCCCAAGTTCCCGAGGACATCACGTGGAGGCGGCCAGGAGATTCT 290
|||||
51 AlaThrProArgLeuPheProAspAspProArgGluGlyGlnVallys 67
|||||
291 GCCACACCCCGGTGTTTCAGATGACCCACGGAAGGCGAGGTGAAGA 340
|||||
67 nGlyLeuGlyAspCysTrpPheLeuCysAlaCysAlaAlaLeuGln 84
|||||
341 GGGCTGCTGGGGATGCTGTGTTCTGTGTGCTCGCGCGCGCTGCAGA 390
|||||
84 ysSerArgHisLeuLeuAspGlnValIleProProGlyGlnProSer 100
|||||
391 AGAGCAGGCACCTCTCGACAGGTCTCTTGCCCTGTGCAG..... 431
|||||
101 AlaaspGlnGluTrpArgGlySerPheThrCysArgIleTrpGlnPhe 117
|||||
432 .....CTTCTCGAGACTGGACTTGCAAAAGTCCAGCTGTA... 467
|||||
117 yArgTrpValGluValThrThrAspAspArgLeuProCysLeu 131
|||||
468 .....TGCTGGAG.....TTCCCATGCTCTG 488

seq_name: /cqn2_5/ptodata/2/1na/6B_COMB.seq:US-09-422-869-17

seq_documentation_block:
; Sequence 17, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: ODA, NAOHISA
; APPLICANT: COX, NANCY J.
; APPLICANT: SREENAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: OTANI, KENICHI
; APPLICANT: HANIS, CRAIG L.
; APPLICANT: BELL, GRAEME I.
; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
; FILE REFERENCE: ARCD:307
; CURRENT APPLICATION NUMBER: US/09/422,869
; CURRENT FILING DATE: 1999-10-21
; EARLIER APPLICATION NUMBER: 60/134,175
; EARLIER FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 864
; TYPE: DNA
; ORGANISM: Human
; OS-09-422-869-17

```

```

US-09-768-877-2 x US-09-422-869-17
Align seg 1/1 to: US-09-422-869-17 from: 1 to: 864

581 GlnValProGluGlyClyArgSerClnhspAlaProProLeuLeuLeuG1 597
|||||
180 CAGGTCCACAGAGGGTGAAGAGAGCCAGGACGACCCACCTGCCTGCCTGCA 229
|||||
597 nGluProLeuLeuSerCysValProHisArgTyrAlaGlnGluValSerA 614
|||||
230 GGAGCGCGTGTGAGCTGCGTCCACATCGCTACGCCAGGAGTGAGCC 279
|||||
614 rGleuCysLeuLeuProAlaglyThrLysValValProSerThrTyr 630
|||||
280 GGCTCTGGCTCTCTGCTCGAGGCACCTAAGAGTGTGGCCCTCCACCTAC 329
|||||
631 LeuProAspThrGluGlyAlaPheThrValThrIleAlaThrArgIleAs 647
|||||
330 CTGCCGGACACAGAGGGGCCCTTCACTGACCATCGCAACACCGAGTGA 379
|||||
647 pArgProSerIleHisSerGlnGluMetLeuGlyClnPheLeuGlnGluV 664
|||||
380 CAGGCGCATCCATTACAGCCAGGAGATGCTGGGCCAGTTCCTCCACAGAG 429
|||||
664 alSerValMetAlaValMetLysThr 672
|||||
430 TCTCGTCTATGGCAGTGATGAAACC 455
|||||

seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-928-692-16
seq_documentation_block:
; Sequence 16, Application US/08928692
; Patent No. 5958727
; GENERAL INFORMATION:
; APPLICANT: Brody, Howard
; APPLICANT: Yaver, Deborah S.
; APPLICANT: Lamsa, Michael
; APPLICANT: Hansen, Kim
; TITLE OF INVENTION: Methods for Modifying the Production of
; NUMBER OF INVENTION: a polypeptide
; NUMBER OF SEQUENCES: 80
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5958727o No. 5958727disk of No. 5958727th.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,692
; FILING DATE: 12-SEPT-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4944.200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4700 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-928-692-16

```

alignment_scores:

Quality: 263.00 Length: 677
 Ratio: 0.843 Gaps: 32
 Percent Similarity: 46.086 Percent Identity: 21.861

alignment block:

US-09-768-877-2 x US-08-928-692-16 ..

Align seg 1/1 to: US-08-928-692-16 from: 1 to: 4700

```

43 TrpArgProGlnGluIleCysAlaThr.....Pr 53
|||||
1944 TGAAGACGACCTTCGGAGCTTCGCGAAGACAGACGAGATGTGACAC 1993
53 oArgLeuPhePro.....AspAspProArgGluGluValIleGlnG 68
|||||
1994 AAGGTGGTCTCTGTGATGCTCTCCAGAAAGACAGATCTGTCCAGG 2043
68 IyLeuLeuGlyAspCysTirPhe.....LeuCysAlaCysAlaIa 81
|||||
2044 ATATGTGACGGACTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 2093
82 ...LeuGlnIySerArgHisLeuLeuAspGlnValIleProGluG 97
|||||
2094 ATGCTGAACCGCGCCAGTGTACTGTAGACAGATGTATCCCTCGCG 2143
97 nProSerTrp..... 100
|||||
2144 ACCGTGATGTGCTGCTGTGACTAATAGTGTAGCATTTTCTCCATG 2193
101 .....AlaAspGlnIuTyArg..GlySerPheThrC 111
|||||
2194 ATACCTAGCCGGGGAGCTCTCAGCTTCACCGCTCAGCAGATATAT 2243
111 yArgGileTrpGlnPheGlyArgTyValGluValThrThrAspArg 127
|||||
2244 TTCGCTTTATTTCAATGGGTCTTCGGAAGTCACTTATGACGACGT 2293
128 LeuProCysLeuAlaGlyArgLeuCysPheSerArgCysGlnArgL 143
|||||
2294 TTGCATGCG.....TCTAAGACATCAAGATCACT 2322
144 .....AspValPheTrpLeuProLeuLeuG 152
|||||
2323 CCACGTGATCGACCGAATAATCCCAATTCCTTGGCGCGCGCTGAG 2372
152 IuTyValTyValAlaTyValHisGlySerTyGlnHisLeuThrAla 168
|||||
2373 AGAAGCGCATTTGAATTTGCGCGAGGCTATGATTTTCCGGAAGCAT 2422
169 GlnValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGlu 185
|||||
2423 TCCGGAGACAGATCTGGGTG...CTGACAGGTTCGATTCGCGACAG 2469
185 PAsnLeuGlyValAlaGlySerGlyGlnGlnAspArgProGlyA 202
|||||
2470 CTTTCTCATATGACGATGTGACGTGCGGACCG.....C 2504
202 rGTrpGlnHisArgThrCysArgGlnLeuLeuHisLeuLysAspG 218
|||||
2505 TCTGGAG...CGACTTACAGATCC...TTTCACCAAGAGATGTCTC 2548
219 Leu...IleSerCysCysValLeuSerProArgAlaGlyAlaArg 234
|||||
2549 TTGACTATAGTACCGGTGACTCAGTGAAGGAA...CAAGAGAACT 2595
234 uGly.....GluPheHisAlaPheIleValSerAspLeuArgGlu 249
|||||
2596 AGGCTCTGTGAGCATGTATGTATCTGATATCAATATGAAGAACT 2645
249 IuGlyGlnAlaGlyGlnCysIleLeuLeuLeuArgGlyIleGlnAs 265
|||||

```

```

2646 AAGTGCAGCA.....CAATTACTCGTGAATAAACCTTGG 2680
266 GlyArgArgCysTrpGlnGlyLeuTrpArgGluGlyGlyGlyTrp 282
|||||
2681 GCT.....GGACGAGATACTGCCCCCGGACAAATGGAAG 2715
282 rGlnValAspAlaAlaValAlaSerGluLeuLeuSerGlnGlnGln 299
|||||
2716 CCTCTGCAATCCAGAGATTTACCCATTAACCCCTCATTTGACCCG 2765
299 IyGluPheTrpValGluGluGluGluPheLeuArgGluPheAspGlu 315
|||||
2766 GTACCTTTTGATGATGTGGAAAGCTCTTCAACATTTGAAACCTC 2815
316 ThrValGlyTyProValThrGluAlaGlyHisLeuGlnSerLeu 332
|||||
2816 TATTTGAAT.....TGAA 2829
332 rGluArgLeuLeuCysHisThrArgAlaLeuProGlyAlaTrp.... 347
|||||
2830 CCGTGAATTTTCAATACCGGAGACGCTCCACTTACGTGGAGCTCA 2879
347 AlIyGlyGlnSerAlaGlyGlyCysArgAsnAsnSerGlyPhePro 363
|||||
2880 ACAACGGAGAGGTGTACCGGCTGT.....TTGTGAT 2914
364 AsnProLysPheTrpLeuArgValSerGluProSerGluValTyLe 380
|||||
2915 AACCGGAGTGTGCAAGTGTCAACCGAGAACGGTGGATGTCTGTG 2962
380 aValLeuGlnArgSerArgLeuHisAlaAlaAspTrpAlaGlyAla 397
|||||
2962 ..... 2962
397 rGAlaLeuValGlyAspSerHisThrSerTrpSerProAlaSerIle 413
|||||
2963 .....CTTCTAGGCAAGCATTTCAAGACAAACAGGAGCGGAGAC 3007
414 GlyLysHisTyGlnAla.....ValGlyLeuHis 424
|||||
3008 CTTGACGAAATACCAAGCAAGTAGAGTCCGCTTTATTAACATAT 3057
424 uTrpLysValGluTyArgArgValAsnLeuProArgValLeuSerMet 441
|||||
3058 CTTTAACCGAGATGGCAACAGGCTCTCTTGTGCT..... 3091
441 rOProValAlaGlyThrAlaCysHisAlaTyArgAspArgGlnVal 456
|||||
3092 .....GATGGGCGCTCACTACATGCT 3109
457 .....LeuArgCysGluLeuSerPr 463
|||||
3110 GGGCCCTATGTGGATTCCTTAATACGCTCATGAGTGAAGAGTCCC 3159
463 O.....GlyTyTyTyLeuAlaValProSerThrPheLeuLysAla 478
|||||
3160 CAGAACAAATACACAGTGTGTCTCCGACAAATCACTGCCATCTTGA 3209
478 rGlyGluPheLeuLeuArgValPheSerThrGlyArgValSerLeu 494
|||||
3210 ATCAAACTTACTTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 3259
495 AlaIleArg.....AlaValAlaLysAsnThrThrProGlyAl 507
|||||
3260 AAGCCCAAGATTAATACATGTGTCTCAGAG..... 3292
507 aAlaLeuProAlaGlyGluTrpGlyThrValGlnLeuArgGlySer 524
|||||
3293 .....ATTCAAGGCTTGG 3307
524 rValGlyGlnThrAlaGlyGlySerArgAsnPheAlaSerTyProThr 540
|||||
3308 ..ACACCTTGACGCGAGGTGGAAATGCCGAATCTTCTCATATTCAC 3355

```


541 AsnProCysPheProPheSerValProGluGlyProGlyPro..... 554
||||| ||| :::::::::: ::
3356 AACCCCAATTAGTTGGAGATAGAGAAATGACACAGATGTTTCACTCCT 3405
555AtgCysValArgIleThrLeuHisGlnHisCysArgProSerAspt 570
::: ||| ::
3406 GCTGGAAATGCCCAACACGGAACCTCCGACCCATGTTAAG..... 3445
570 hrGluPheHisProIleGlyPheHisIlePheGlnValProGluGlyGly 586
||| ||| :::::::::: ::
3446 ..TTATTCGTGTCCTCAATGGAATCGTGTCTCGGAGTA..... 3481
587 ArgSerGlnAspAlaProProLeuLeuLeuGlnGluProLeuLeuSerCy 603
||||| ::
3482 CGCAGTCGGACATA.....ATCGCTGATAG 3507
603 sValProHisArgTyrAlaGlnGluValSerArgLeuCysLeuLeuProA 620
::: ||| ::
3508 TGGTGACTATCGCGGTGGTGGCTCCCTGTGGAAAGAGGCTCTCGAAG 3557
620 laGlyThrTyrLysValValProSerThrTyrLeuProAspThrGluGly 636
||: ||| ::
3558 CGGGCTCATATCAATCGTCTGTCCACATTCGCGCGGATCACTTGGC 3607
637 AlaPheThrValThrIleAlaThrArgIle 646
||||| ::
3608 CGATTCACGCTCTGGGTATCTCTCTTAGTT 3637

Page 1

About: Results were produced by the Gencore software, version 4.5
Copyright (c) 1993-2000 Compugen Ltd.

```

-MODEL=frame+g2n.model -DEV=xip
-0/cgcn2.1/USPTO.spool.US9768877/runat.16092002.15.811.8332/app.query.fasta.1.733
-DB=EST -OEMT=fastap -SUFFIX=fast -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEC=0.000 -LOOPEXT=0.000 -GAPOP=4.500
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=biyosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=10.00 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUEMT=pfis -NORM=ext -HEADSIZE=500
-MINLEN=0 -MAXLEN=20000000000 -USR=US9768877_cgcn2.1.5424
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMECUT=120 -WARN_TIMECUT=30
-NO_XLPHY -WAIT -THREADS=1

```

Database length: -1841457050

Sequence	Strd	Orig	Zscore	EScore	len	Document
gb_eest1:AL526645	+	1413.00	2053.36	4.0e-105	908	AL526645 ALT26645 LTI_NFL003 NE
gb_eest2:BI916764	+	1304.00	1894.03	3.0e-96	887	BI916764 6031786795F1 NIH_MGC_12
gb_eest2:BG147596	+	1241.50	1803.13	3.5e-91	842	BG147596 60252151077F1 NIH_MGC_2
gb_eest2:BM477401	+	1170.00	1696.45	3.0e-85	999	BM477401 AGENCOURT_648488 NIH_N
gb_eest2:BI1666734	+	1163.00	1669.72	7.2e-85	738	BI1666734 603293113F1 NIH_MGC_9E
gb_eest1:AL529373	+	1165.00	1667.51	9.5e-85	787	AL529373 AL529373 LTI_NFL001 NE
gb_eest2:BE889733	+	1115.00	1618.97	6.3e-81	769	BE889733 601512755F1 NIH_MGC_77
gb_eest2:BF313048	+	1106.00	1603.43	4.6e-80	942	BF313048 601895051F1 NIH_MGC_19
gb_hlc:AK002548	+	1091.00	1579.28	1.0e-78	1137	AK002548 MUs musculi adult ma
gb_eest2:BI1410486	+	1091.00	1578.01	1.2e-78	1269	AK013497 MUs musculi adult ma
gb_eest2:BE3390603	+	1078.00	1551.42	1.0e-77	1028	BI1410486 602963161F1 NCI_CGAP
gb_eest2:BI1412050	+	1076.50	1554.73	5.6e-78	601	BE3390603 601284961F1 NIH_MGC_44
gb_eest2:BE6193302	+	1066.50	1548.55	5.2e-77	940	BI1412050 602965960F1 NCI_CGAP_L
gb_eest1:AL558905	+	1056.50	1533.01	3.8e-76	789	BE6193302 602965960F1 NCI_CGAP_L
gb_eest2:BE533889	+	1047.00	1519.30	2.2e-75	776	BE6193302 601473184F1 NIH_MGC_68
gb_eest2:BI1108473	+	1043.00	1512.11	5.6e-75	870	AL558905 LTI_NFL008 20C
gb_eest2:BE294553	+	1014.50	1470.72	1.1e-72	845	BI108473 602460221F1 NIH_MGC_20
gb_eest2:BE293992	+	1012.00	1469.38	1.3e-72	692	BE294553 6024694921F1 NCI_CGAP_N
gb_eest1:AL526168	+	995.00	1441.57	5.0e-71	715	BE293992 601173172F1 NIH_MGC_17
gb_eest2:BG137387	+	975.00	1415.18	1.3e-69	671	AL526168 ALT26168 LTI_NFL003 NE
gb_eest2:BI913176	+	973.00	1406.54	4.2e-69	1134	BG137387 60233841F1 NCI_CGAP
gb_eest2:BI1545057	+	970.00	1408.44	3.3e-69	660	BI915176 603317733F1 NIH_MGC_9E
gb_eest2:BF027430	+	966.00	1401.05	8.6e-66	753	BI1545057 603242542F1 NIH_MGC_19
gb_eest2:BF025700	+	949.00	1380.04	1.3e-67	539	BF027433 6016742551F1 NIH_MGC_20
gb_eest2:BE261943	+	927.00	1339.21	2.4e-65	1132	BE025700 601670655F1 NIH_MGC_2
gb_eest2:AL1582661	+	925.50	1345.42	1.1e-65	549	BE261943 601147487F1 NIH_MGC_19
gb_eest2:BF301261	+	925.00	1342.12	1.6e-65	685	BI188804 6030321408F1 NIH_MGC_11
gb_eest1:AL526618	+	919.00	1329.72	8.0e-65	929	BF301261 6020291073F1 NCI_CGAP_S
gb_eest2:BE3702088	-	915.50	1323.96	1.7e-64	987	AL526618 ALT26618 LTI_NFL003 NE
gb_eest2:BF166926	+	879.00	1274.98	9.2e-62	678	BE3702088 602683618F1 NIH_MGC_95
gb_eest1:AL526112	+	849.00	1226.81	4.4e-59	969	BF166926 601774074F1 NCI_CGAP_L
gb_eest2:BF236646	+	818.50	1182.81	1.2e-56	916	AL526112 ALT26112 LTI_NFL003 NE
gb_eest2:BE8883220	+	802.00	1161.17	2.0e-55	717	BF236646 602028634F1 NCI_CGAP_L
gb_eest2:BE788676	+	797.00	1155.71	4.0e-55	938	BE8883220 601508648F1 NIH_MGC_68
gb_eest2:BE3390299	+	786.00	1135.05	5.6e-54	920	BE788676 601477556F1 NIH_MGC_77
gb_eest2:BE33910796	+	767.50	1109.72	1.4e-52	799	BE3390299 602185192F1 NIH_MGC_44
gb_eest2:BF3130129	+	756.00	1090.52	1.7e-51	979	BF310796 601895608F1 NIH_MGC_19
gb_eest2:BE312136	+	730.00	1058.18	1.1e-49	916	BE312136 6026297103F1 NIH_MGC_97
gb_eest1:AL563429	-	722.50	1046.00	5.1e-48	945	BE312136 6011454948F1 NIH_MGC_19
gb_eest1:AL563429	-	715.50	1031.51	3.3e-48	955	AL563429 ALT563429 LTI_NFL001 NE

2019年12月10日

51 ALATHrProARgLeuPheProASpSPProARgGLuGLyGlnValLysG1 67

```
|||||
274 GCCACACCCCGGCTGTTTCCAGTACACCCACGGAAGGCGAGTGAAGCA 323
67 nGlyLeuLeuGlyAspCysTrpPheLeuCysAlaCysAlaLeuGlnL 84
|||||
324 GGGGCTGCTGGGGGATGCTGTTTCTGTGCTGCTGCGCGCGTGCAGA 373
84 ySerArgHisLeuLeuAspGlnValIleProProGlyGlnProSerTrp 100
|||||
374 AGAGCAGGCACCTCTCGGACCAAGTCACTTCTCCGGGACAGCGAGTGG 423
101 AlaAspGlnGluTyrArgGlySerPheThrCysArgIleTrpGlnPheG 117
|||||
424 GCGCACCAGGAGTACCGGGGCTCTTCACTCTGCTGCTGCGAGTTGG 473
117 YArgTrpValGluValThrAspAspArgLeuProCysLeuAlaGlyA 134
|||||
474 AGCGTGGGTGGAGTGACACAGATGACCGCTGCGTGCCTTGCAGGGA 523
134 rGLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLe 150
|||||
524 GACTCTGTTCTCCCGCTGCCAGAGGAGGATGTGTTCTGCTGCCCTTA 573
151 LeuGluLysValTyrAlaLysValHisGlySerTyrGluHisLeuTrp 167
|||||
574 CTGGAAAGGCTACGCCAAGGTCCATGGTCTTACGAGCACCTGTGGGC 623
167 aGlyGlnValAlaAspAlaLeu.ValAspLeuThrGlyGlyLeuAlaGlu 183
|||||
624 CCGGCAGGTGGGATGTCCTGGTGGACCTGACCGCGGCGCTGGCAGAA 673
184 ArgTrpAsnLeuLysGlyValAlaGlySerGlyGlyGlnGlnAspArg 200
|||||
674 AGATGAACCTCAAGGGGCTASAGGAGGAGGAGGAGGAGGAGGAGGAG 723
200 oGlyArgTrpGluHisArgThrCysArgGlnLeuLeuHisLeuLysAsp 217
|||||
724 AGCGCGCTGGGAGCAGAGCTTGTGGCAGCTGTCCACCTGAAGGACC 773
217 InCysLeuIleSerCysValLeuSerProArgAlaGlyAlaArgGlu 233
|||||
774 AGTGCTGATCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 823
234 LeuGlyGluPheHisAlaPheIleValSerAspLeuArgGluGlnG 250
|||||
824 CTGGGGGAGTTCATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 873
250 yGlnAlaGlyGlnCysIleLeuLeuLeuArgIle 261
|||||
874 TCAGCGGGGCGAGTGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 907
```

seq_name: gb_est2:B1916704

```
seq_documentation_block:
LOCUS      B1916704                887 bp    mRNA    linear    EST 16-OCT-2001
DEFINITION 603178679F1 NIH_MGC_121 Homo sapiens cDNA clone IMAGE:5242963 5',
            mRNA sequence.
ACCESSION  B1916704
VERSION    B1916704.1 GI:16180666
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 887)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
```

DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM11612 row: n column: 20
High quality sequence stop: 809.

FEATURES
source
1..887
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5242963"
/clone_lib="NIH_MGC_121"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: EcoRV (destroyed); RNA source anonymous pool of 3
fetal brains, female age 20 weeks, female age 24 weeks,
and male age 26 weeks. Library is oligo-dT primed and
directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.7 kb, insert size range
0.7-3.5 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 017. Note:
this is a NIH_MGC Library."

BASE COUNT 163 a 260 c 310 g 154 t
ORIGIN

alignment_scores:
Quality: 1304.00 Length: 299
Ratio: 4.559 Gaps: 8
Percent Similarity: 95.652 Percent Identity: 92.308
alignment_block:
US-09-768-877-2 x B1916704

Align seg 1/1 to: B1916704 from: 1 to: 887

```
211 LeuLeuHisLeuLysAspGlnCysLeuIleSerCysValLeuSerPr 227
|||||
2 CTGCTCCACCTGAAGGACCACTGCTGATCAGCTGCTGCTGCTGCTGCTG 51
|||||
227 oArgAlaGlyAlaArgGluLeuGlyGluPheHisAlaPheIleValSerA 244
|||||
52 CAGAGCAGGTGCGCGGAGCTGGGGAGTCCATGCCCTTCATTGCTCGG 101
|||||
244 sPLeuArgGluLeuGlnGlyGlnAlaGlyGlnCysIleLeuLeuLeuArg 260
|||||
102 ACCTGCGGGAGCTCCAGGCTCAGCGGCGCCAGTGCATCTGCTGCTGCGG 151
|||||
261 IleGlnAsnProTrpGlyArg.ArgCysTrpGlnGlyLeuTrpArgGluG 277
|||||
152 ATCCAGAACCCCTGGGCGCTGGCGGTGCTGGCAGGGGCTCTGGAGAGAG 201
|||||
277 lyGlyGluGlyTrpSerGlnValAspAlaAlaValAlaSerGluLeuLeu 293
|||||
202 GGGGTGAAGGTGAGCAGCAGGTAGATGACGCGGTAGCATCTGAGCTCCTG 251
|||||
294 SerGlnLeuGlnGluGlyGluPheTrpValGluGluGluPheLeuArg 310
|||||
252 TCCAGAGTCCAGGAAGGGGAGTCTCTGGGTGGAGGAGGAGGAGTTCCTC 301
|||||
310 gGluPheAspGluLeuThrValGlyTyrProValThrGluAlaGlyHisL 327
|||||
302 GGAGTTTGACGAGCTACCGTTGGCTACCGCGGTACCGAGGCGCGGCCACC 351
|||||
327 euGlnSerLeuTyrThrGluArgLeuLeuCysHisThrArgAlaLeuPro 343
|||||
352 TGCAGAGCCTTACACAGAGAGGCTGCTCTGCCATACGCGGGCGCTGCT 401
|||||
344 GlyAlaTrpValLysGlyGlnSerAlaGlyCysArgAsnSerG 360
|||||
402 GGGGCGCTGGGTCAAGGGGCCAGTTCAGCAGGAGGCTGCCGGAACAACAG 451
```

```

360 yPheProSerAsnProLysPheTripleuArgValSerGluProSerGluV 377
452 CTTTCCACCAACCCCAATTTCTGGCTGGGGCTCAGAACCGATGAGG 501
377 aLysr1LeaValLeuGlnArgSerArgLeuHisAlaAlaAspTrpAla 393
502 TGTACATTCCTGCTCCAGAGATTCAGAGCTGACCGCGGAGCTGGGA 551
394 GtArgAlaArgAlaLeuValGlyAspSerHisThrSerTrpSer. ProA 410
552 GGGCCGGCC. CGGGCACTGGTGGGTGACACTGACTTCTGAGGCCCCAG 600
410 lAser1LeProGlyLysHisTyrGlnAlaValGlyLeuHisLeuTrpLys 426
601 CGAGCATCCCGGGCAAGCACTACAGAGCTGTGGGTCTGACCTGTGGAG 650
427 .ValGluLysArgArgVal. AsnLeuProArg. ValLeuSerMetProPr 442
651 GGTAGAGAAGCGCGGCTCAATCTGCCTAGGGGTCTCTTCATGCCCCC 700
442 oValAlaGlyThrAlaCysHisAlaTyrAspArgGluValHisLeuArgC 459
701 CGTGCTGGCACCGGTGCCAATGATACAGACCGGAGGTCCACTGCGGT 750
459 ySGluLeuSerProGlyTyrTyrLeuAlaValProSerThrPheLeuLys 475
751 GTGACCTTCACG. GGTACTACTCTGAGCTGTCCAG. AACTTCTGAG 798
476 AspAlaProGlyGluPheLeuLeuArgVal. PheSerThrGlyArg. Val 491
799 GACCGCGGAGGAAGTCTGCTCGAGCTTACTACTGCGGAGAACTC 848
492 SerLeuSerAlaLeuArgAlaValAlaLys 501
849 TCCCTTAGCGGCATCCAGGAGTGGCAAGA 879

seq_name: gb_est2:BG475966

seq documentation block:
LOCUS BG475966 842 bp mRNA linear EST 21-MAR-2001
DEFINITION 602521077F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:4639579 5',
ACCESSION BG475966
VERSION BG475966.1 GI:13408245
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 842)
NIH-MGC http://mhc.ncl.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC/DCPD/DPF
cDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHCMI403 row: 1 column: 20
High quality sequence stop: 760.
Location/Qualifiers
1..842
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:4639579"
/clone_lib="NIH_MGC_20"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pORF7; Site_1: XhoI; Site_2:"

```

ECORI; cDNA made by oligo-dT priming. Directionally cloned into EcorI/XhoI sites using the following 5' adaptor: GGCAGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies).*

BASE COUNT
ORIGIN
164 a 238 c 307 g 133 t

alignment_scores:
Quality: 1241.50 Length: 282
Ratio: 4.721 Gaps: 11
Percent Similarity: 93.262 Percent Identity: 91.135

alignment_block:
us-09-768-877-2 x BG475966

Align seg 1/1 to: BG475966 from: 1 to: 842

```

193 SerGlyGlyGlnGlnAspArgProGlyArgTyrGlnHisArgThrCysAr 209
|||||
2 AGCGGAGGCCAGACAGACAGGCGCGCTGGGAGACAGACTGTGCG 51
209 gGlnLeuLeuHisLeuLysAspGlnCysLeuLeuSerCysValLeuS 226
52 GCAGCTGCTCCACCTGAAGACCACTGTCTGATCAGCTGCGCTCA 101
226 erProArgAlaGlyAlaArgGluLeuGlyGluPheHisAlaPheLeuAl 242
102 GCGCCAGAGCAGTCCCGGAGCTGGGGAGTTCATGCTTATTTTC 151
243 SerAspLeuArgGluLeuGlnGlyGlnAlaGlyGlnCysIleLeuLeu 259
|||||
152 TCGGACCTGCGGAGACTCCAGGCTCAGGCGGCGGCATCTCTGCT 201
259 uArgIleGlnAsnProTyrGlyArgCysTyrPdlGlnIleuTrpArg 276
202 GCGATCCAGAAACCCCTGGGCGCGGCTGCTGCGAGGCGCTGGAAG 251
276 LngIlyGlyGlnGlyTyrSerGlnValAspAla. AlayaAlaSerGlu 292
252 AGGGGGGGAAGGCTGAGCCAGGTAGATCAAGCGGTAGCATCTGACT 301
292 uLeuSerGlnLeuGlnGlyGluGlyGluPheTrpValGluGluGluPhe 309
302 CTTCTCCAGCTCCAGAAAGGAGTCTGGTGGAGAGAGAGAGTTC 351
309 euArgGluPheAspGluLeuThrValGlyTyrProValThrGluAlaGly 325
352 TCAGGAGGTGGAGAGAGCTACCTAGGCTACCGGTCACGAGCGCGC 401
326 HisLeuGln. SerLeuTyrThrGluArgLeuLeuCysHisThrArgAla 342
402 CACTTCAGAGAGCTCTACACAGAGAGCTCTGCGCATACGGGGGGC 451
342 euProGlyAlaTrpValGlyGlnSer. AlaGlyGly. CysArgAsn 358
452 TGCCTGGGCTGGGTCAAGGCCAGTCCAGCAGAGAGCTGACCGGACA 501
358 snSerGlyPheProSer. AsnProLysPheTripleuArgValSerGlu 374
|||||
502 ACAGCGGCTTACCCCAAGAAACCCCAATTTCTGCTCGGCTCTCAAC 551
374 roSerGluValTyrIleAlaValLeuGlnArgSerArgLeuHisAlaAla 390
552 CGAGTGGGTGTACATGCGCTCTGACAGATCCAGAGCTGACGCGGCA 601
391 .AspTrpAlaGlyArg. AlaArgAlaLeuValGlyAspSerHisThrSer 406
|||||
602 GGACTGGGAGCGCGGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 651

```

407 TrpSerProAlaSerIleProGlyLysHisTyrGlnAlaValGlyLeuH 423
||||| ||||||| ||||||| ||||||| ||| :||
652 TGGAGCAGCAGAGCATCCGGCGAAGCACTACAGGCTGAGGTCATGC 701
423 isLeuTrpLysValGluLysArgArgValAsnLeuProArgValLeuSe 439
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
702 ACCTCTGGAGGTAGAGAGCGAGGGTCAATCTCCCTAGGCTCTGTC 751
439 rMetProValAlaGlyThrAlaCysHisAlaTyrAspArgGluVal 455
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
752 AATGCCCGCTGGCTGGCACCGGTG...CATGCATACGACCGGAGGTA 798
456 HisLeuArgCysGluLeuSerProGly 464
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
799 CACCTGGTAGTGAGCTCACCGGGT 825

seq_name: gb_est2:BM477401

seq_documentation_block:
LOCUS BM477401 999 bp mRNA linear EST 05-FEB-2002
DEFINITION AGENCOURT_648488 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:5554201
5', mRNA sequence.

ACCESSION BM477401
VERSION BM477401.1 GI:18526443
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 999)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Lou Staudt
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLML)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLML at:
<http://image.llnl.gov>

Plate: L1AM12272 row: g column: 02
High quality sequence stop: 570.
Location/Qualifiers
1..999
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="5554201"
/clone_lib="NIH_MGC_85"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lymph. Vector: pCMV-SPORT6; Site_1: NOTI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.867 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 148 a 319 c 382 g 150 t

FEATURES

source
1..999
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="5554201"
/clone_lib="NIH_MGC_85"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lymph. Vector: pCMV-SPORT6; Site_1: NOTI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.867 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 148 a 319 c 382 g 150 t

alignment_scores:

Quality: 1170.00 Length: 307
Ratio: 4.661 Gaps: 9
Percent Similarity: 81.759 Percent Identity: 76.873

alignment_block:

US-09-768-877-2 x BM477401 ..
Align seg 1/1 to: BM477401 from: 1 to: 999
1 MetArgAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||

134 ATCGCGGGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 183
17 aaLpheProAlaAlaAspSerSerLeuPheCysaspLeuSerThrProL 34
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
184 CGCCTTCCCGCGCGGAGACTCTCGCTCTCTGCGACTTGTCTACGCCGC 233
34 euAlaGlnPheArgGluAspIleThrTrpArgArgProGlnGluIleCys 50
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
234 TGGCCAGTTCCGCGGAGGACATCAGTGGAGGCGGCGGCGGCGGAGATTGT 283
51 AlaThrProArgLeuPheProaspaspProArgGluGlyGlnValLysG 67
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
284 GCCACACCCCGGCTGTTCAGATGACCCACGGAAGGCGGCGGAGATTGT 333
67 nGlyLeuLeuGlyAspCysTrpPheLeuCysAlaCysAlaAlaLeuGlnL 84
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
334 GGGGCTGCTGGGGGATGCTGTTCCTGTGCTGCGCGCGGCTGCAGA 383
84 ysSerArgHisLeuLeuaspGlnValIleProProGlyGlnProSerTrp 100
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
384 AGAGCAGGCACCTCTCGGACAGGTCAATTCCTCCGGGACAGCGGAGCTGG 433
101 AlaaspGlnGluTyrArgGlySerPheThrCysArgIleTrpGlnPheG 117
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
434 GCCGACCGAGGAGTACCGGGGCTCCTTCACCTGCGCATTTGCGAGTTTGG 483
117 yArgTrpValGluValThrThrAspAspArgLeuProCysLeuAlaGlyA 134
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
484 AGCTGGGTGGAGGTGACACAGATGACCGCTGCGCTGCGCTTGCAGGGA 533
134 rGLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
534 GACTCTGTTCTCCCGCGGCGGAGAGGAGTGTCTTGGCTCCCTTFA 583
151 LeuGluLysValTyrAlaLysValHisGlySerTyrGluHisLeuTrpAl 167
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
584 CTGGAAGAGGTCTACGCCAAGGTCCATGGGTCTTACGAGCAGCTGTGGGC 633
167 aGlyGlnValAlaAspAlaLeuValaspLeuThrGlyGlyLeuAlaGluA 184
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
634 CGGGCAGGTGGCGGATGCGCTGCGGCTGACCGGCGGCGCTGGCAGAAA 683
184 rGTrpAsnLeuLysGlyValAlaGlySerGlyGlnGlnAspArgPr 200
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
684 GATGAACCTGGAAGGGGTAGCAGGAAGCGGAGGCGGCGGAGCAGAGGCC 733
200 oGlyArgTrpGluHisArgThrCysArgGlnLeuLeuHisLeuLysAs 216
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
734 AGGCGCTGGGAGACACCGGACTTGTGCGCGGCTGCTCCACTGAAA... 781
216 pGlnCysLeuIleSerCysValLeuSerProArgAlaGlyAlaArg. 232
782GGACCGCGGTCTGGTCCCACTG 803
233GluLeuGlyGluPhe 237
804 CGGCGGCTCCACCCCAACAAGATGCCCGGGAAGCTGGGGAGCTC 853
238 HisAlaPheIleValSerAspLeuArgGlu.....LeuGlnG 250
||||| ||||||| |||||||
854 CTGCGCTTTTGTCTCCCGGAGCGCGCGGAAACTTCCAGGGGTTCCAGGG 903
250 yGlnAlaGlyGlnCysIleLeuLeuLeuArgIleGlnAsnProTrpGly. 266
||||| ||||||| |||||||
904 GGGCGCGGAGCTCCCGGCTGCTGGGGAATCCCAACACCCCTGGGAGC 953
267 ArgArgCysTrpGlnGlyLeuTrpArgGluGlyGlyGlyTrpSerG 283
||||| ||||||| |||||||
954 CGGCGG.....GGCGCATCCGGGGGGCCCCACTGCGCACC 988
283 nValAspAlaAla 287
: |||||||
989 G....GACGCGCG 998

seq_name: gb_est2:B1668734

seq_documentation_block:

LOCUS B1668734 738 bp mRNA linear EST 12-SEP-2001

DEFINITION 603293173F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5312178 5',

ACCESSION B1668734

VERSION B1668734.1 GI:15582967

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo. 1 (bases 1 to 738)

AUTHORS NIH-MGC <http://mgs.ncl.nih.gov/>

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Miklos Palokovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki Toshiyuki and Piero Carninci (RIKEN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <http://image.linl.gov>
Plate: L1AM11791 row: b column: 19

High quality sequence stop: 738.

FEATURES

source

1..738

Location/Qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:5312178"

/clone_1lb="NIH_MGC_96"

/issue_type="hypothalamus"

/lab_host="DH10B"

/note="Organ: brain; Vector: pBluescript (modified pBluescript KS+); Site: 1: BamHI; Site 2: SalI-XhoI (gtcgag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTAA-3', size-selected for average insert size 2.3 kb and normalized to 10⁷ 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT 120 a 255 c 237 g 126 t

ORIGIN

alignment_scores:

Quality: 1163.00 Length: 243

Ratio: 5.035 Gaps: 0

Percent Similarity: 95.062 Percent Identity: 93.004

alignment_block:

us-09-768-877-2 x B1668734 ..

Align seg 1/1 to: B1668734 from: 1 to: 738

369 leuArGValSerGIuProSerGIuValTyrTleAlaValleuGlnArgSe 385

6 GTGGGGGTCTCAGAACGAGTGTACATTCCTCCGCGAGATC 55

385 rArgLeuHisAlaAlaAspTrpAlaGlyArGAlaArGAlaLeuValGly 402

56 CAGGCTCCAGCGCGGCGACTGGCGAGCCGCGGCGGCGACTGGTGGTG 105

402 spSerHisThrSerTrpSerProAlaSerIleProGlyGlySHistyrGln 418

106 ACAATCTACTCTGTGGAGCCAGGAGCATCCGCGGCGGCGACTTACAG 155

seq_name: gb_est1:AL529373

seq_documentation_block:

LOCUS AL529373 787 bp mRNA linear EST 13-FEB-2001

DEFINITION AL529373 LYL_NFL001.NBC4 Homo sapiens cDNA clone CS0DD006VAL5 5

ACCESSION AL529373

VERSION AL529373.1 GI:12792866

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo. 1 (bases 1 to 787)

AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polayes, D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished (2001)

COMMENT Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr

FEATURES

source

1..787

Location/Qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

419 AlaValGlyLeuHisLeuTrpIlyValGlyLysArgAlaValAsnLeuPr 435

156 GCTGTGGGTCTGCACCTGTGGAAGTAGAAGCGCGGTCATCTCC 205

435 oArGValLeuSerMetProValAlaGlyThrAlaCysHisAlaTyrA 452

206 TAGGCTCCTGTCCATGCCCCCGTGGTGGCAGCCGTCGATGATC.G 254

452 spArgGlyValHisLeuArgCysGlyLeuSerProGlyTyrTyrLeuAla 468

255 ACCGGAGGTCCACCTGCTGTGAGCTGTCACCGGCTACTACTGGCT 304

469 ValProSerThrPheLeuLysAspAlaProGlyGluPheLeuAlaYva 485

305 GTCCCAACACCTTCTTAAGGAGCGCCAGGAG.TTCTGCTCCGAGT 353

485 IPhSerThrGlyArgValSerLeuSerAlaIleArgAlaValAlaTysa 502

354 CTTCCTACCCGGCGAGTCTCCCTTAGCGCATCGGCACTGGCCAGA 403

502 snThrTrpProGlyAlaAlaLeuProAlaGlyGluTrpGlyThrValGln 518

404 ACACCAACCCCGGGGACGCTGCTGGCGGGGAGTGGGGAGCGTGCGAG 453

519 leuArGlySerTrpArgValGlyGlnThrAlaGlyGlySerArgAsnPh 535

454 CTACGGGGTCTTGGAGGTGCGGACAGCGGGGGGCGACAGAACTT 503

535 eAlaSerTrpProThrAsnProCysPheProPheSerValProGluGlyP 552

504 TGCCTCATACCCCAACCAACCTGCTTCCCTTCGCGGTCGCCAGGGGCC 553

552 roGlyProArgCysValArgIleThrLeuHisGlnHisCysArgProSer 568

554 CTGGCCCCCGGTGGCTCCGACATCTGCATCGCATGCTGGCCCCAGT 603

569 AspThrGluPheHisProIleGlyPheHisIlePheGlnValProGluG 585

604 GACACCGAGTTCACCCCATGCGCTTCATATCTCCAGGTCGCCAGAGGG 653

585 yGlyArgSerGlnAspAlaProProLeuLeuGlnGluProLeuLeuS 602

654 TGAAGAGGAGCGAGAGCACCCACCTGCTGTGCAGAGACCGTGTCTGA 703

602 erCysValProHisArgTyrAlaGlnGlu 611

704 GCTGCTGCCACATCGCTACGCCAGGAG 732

/clone="CS0DD006YA15"
 /clone_lib="LTI_NFL001_NBC4"
 /sex="male"
 /tissue_type="neuroblastoma cells"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"
 BASE COUNT 123 a 237 c 296 g 125 t 6 others
 ORIGIN

alignment_scores:
 Quality: 1162.00 Length: 219
 Ratio: 5.380 Gaps: 0
 Percent Similarity: 98.630 Percent Identity: 98.174

alignment_block:
 US-09-768-877-2 x AL529373

Align seg 1/1 to: AL529373 from: 1 to: 787

1 MetArgAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
 131 ATGCGCGGGCGGGCGGCGGCGGAGGAGCTGTTCGGGAGCGC 180
 17 aAlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrProL 34
 181 CGCCTTCCCGCGCGGACTCTCTCTTCGACATGTGTACGCCGC 230
 34 euAlaGlnPheArgGluAspIleThrArgArgProGlnGluLeuCys 50
 231 TGGCCAGTTCGCGGAGGACATCAGTGGAGCGGCCCGGAGATTGT 280
 51 AlaThrProArgLeuPheProAspAspProArgGluGlyGlnValLysG 67
 281 GCCACACCCCGGCTTTCCAGATGACCCAGGGAAGGCGAGGTGAACA 330
 67 nGlyLeuLeuGlyAspCysTyrPheLeuCysAlaCysAlaAlaLeuGln 84
 331 GGGGCTGCTGGGGATTGCTGCTCTGTGCTGCGCCCGCGCTGCAGA 380
 84 ySerArgHisLeuLeuAspGlnValIleProProGlyGlnProSerTrp 100
 381 AGAGCAGGCACCTCTCGGACCAAGGTCAATCTCGGGACAGCGAGCTG 430
 101 AlaAspGlnGluTyrArgGlySerPheThrCysArgIleTyrGlnPheG 117
 431 CGCGACAGGAGTACCGGGCTCTTACCTGCTCCCAFTTGGCAGTTTG 480
 117 yArgTrpValGluValThrAspAspArgLeuProCysLeuAlaGlyA 134
 481 ACGCTGGGTGGAGGTACACAGATGACCGCTGCCGTGCTTCGACGGA 530
 134 rGluLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuPro 150
 531 GACTCTGTCTTCCTCCGCTGCCAGAGGAGGATGTCTTGGCTCCCTTA 580
 151 LeuGluLysValTyrAlaLysValHisGlySerTyrGluHisLeuTrpAl 167
 581 CTNGAAADGCTACGCCAANGTCCATGGGTCTTACGAGCACCTGTGGGC 630
 167 aGlyGlnValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGluA 184
 631 CGGGCAGGTGGGGATGC CTGGTGGACCTGACCGGGCGGCTGGCAGAA 679

184 rGTrpAsnLeuLysGlyValAlaGlySerGlyGlyGlnGlnAspArgPro 200
 680 GATGAACCTGAAGGCGGTASAGGAGCGGAGCCAGCAGGACAGGCCA 729
 201 GlyArgTrpGluHisArgThrCysArgGlnLeuLeuHisLeuLysAspG 217
 730 GGCCCTGGGAGCACAGGACCTGTTCGGCAGCTGCTCCACCTCAAGGACCA 779
 217 nCysLeu 219
 780 GTGKCTG 786

seq_name: gb_est2:BE889733

seq_documentation_block:

LOCUS BE889733 769 bp mRNA linear EST 20-OCT-2000
 DEFINITION 601512765F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3914009 5',
 mRNA sequence.

ACCESSION BE889733
 VERSION BE889733.1 GI:10347351

KEYWORDS EST.
 SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 769)

NIH-MGC http://mgi.nci.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cyapbs-r@mail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM9735 row: a column: 18

High quality sequence stop: 696.

FEATURES

source

1..769

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3914009"

/clone_lib="NIH_MGC_71"

/tissue_type="leiomyosarcoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;

Site_2: SalI; Cloned unidirectionally. Primer: oligo dt.

Average insert size 2.1 kb.

BASE COUNT 131 a 262 c 238 g 138 t

ORIGIN

alignment_scores:

Quality: 1115.00 Length: 221

Ratio: 5.068 Gaps: 0

Percent Similarity: 99.548 Percent Identity: 99.095

alignment_block:

US-09-768-877-2 x BE889733

Align seg 1/1 to: BE889733 from: 1 to: 769

452 AspArgGluValHisLeuArgCysGluLeuSerProGlyTyrTyrLeuAl 468
 1 GACCGGAGGTCCACCTGCGTTGTGAGCTCTCACCGGGCTACTACCTGGC 50

468 aValProSerThrPheLeuLysAspAlaProGlyGluPheLeuLeuArgV 485
 51 TGTCCCCAGCACCTTCTCTGAAGGATGCCAGGGGAGTTCTCTCTCCGAG 100

485 alpHeserThrglyArvalSerleuSerAlaIleArAlaValAlaLys 501
 101 TCTTCTTACCGGGGAGTCTCTTACGGCCATCAGGCGACATGGCCAG 150
 502 AsnThrThrProGlyAlaAlaIleuProAlaGlyIleuTrpGlyThrValG1 518
 151 AACACCGCCCCGGGGACACCTGCTGGGGAGTGGGGAGCCGTGCA 200
 518 nLeuArglySerTrpArgValAlGlyIleThrAlaGlySerArgAsp 535
 201 GCTACGGGGTCTTGAGAGTGGCCAGACGGCGGGGGACACAGAACT 250
 535 hAlaSerTrpThrAsnProCysPheProPheSerValProGly 551
 251 TTGCTCATACCCACCAACCCCTCTCCCTCTCGGTCGCCGAGGC 300
 552 ProGlyProArgCysValArgIleThrIleuHisGlnHisCysArgProse 568
 301 CTTGGCCCCCGTGGCTGGCATCTGTGACATGACATGCGGCCAG 350
 568 rAspThrGluPheHisProIleGlyPheHisIlePheGlnValProGly 585
 351 TGACACCGAGTTCACACCCATCGGCTTCATATCTCCAGTCCAGAG 400
 585 IyGlyArgSerGlnAspAlaProProIleuLeuGlnIleuProLeu 601
 401 GTGAGAGGAGCGACGACCCACCCACCTGCTGACAGAGCCGCTGCTG 450
 602 SerCysValProHisArgTrpAlaGlnIleuValSerArgLeuCysLeu 618
 451 AGCTCGGCGCCATCGCTC. GCCAGAGAGTGAGCCGCGCTGCTCT 499
 618 uProAlaGlyThrTrpValValProSerThrTrpLeuProAspThr 635
 500 GCTGCGGGACACTACAGAGTGTGCTCCCTCCACTACCTGCGGAGACAG 549
 635 IuGlyAlaPheThrValThrIleAlaThrArgIleAspArgProSerIle 651
 550 AGGGGGCTTACAGTGCATCGCAACCCAGATTGACAGGCCATCCATT 599
 652 HisSerGlnIleuLeuGlyIlePheLeuGlnIleuValSerValMetAl 668
 600 CACAG. CAGGAGATGCTGGCCAGTCTCCACAGAGGTCTCCGTCATGGC 648
 668 aValMetLysThr 672
 649 AGTCATGAAAAC 661

seq_name: gb_est2:BF310408

seq_documentation_block:

LOCUS BF310408 942 bp mRNA linear EST 21-NOV-2000
 DEFINITION 601895061F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4124630 5',
 mRNA sequence.
 ACCESSION BF310408
 VERSION BF310408.1 GI:11257971
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 942)
 NIH-MGC http://mhc.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapds-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov

FEATURES
 Source
 Plate: LHCN1011 row: 1 column: 15
 High quality sequence stop: 718.
 Location/Qualifiers
 1. 942

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4124630"
 /clone_id="NIH_MGC_19"
 /tissue_type="neuroblastoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: brain; Vector: pORF7; Site:1: XhoI; Site:2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCAAGAG(G). Library constructed by Ling Hong
 in the laboratory of Gerald W. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC library."

BASE COUNT 193 a 250 c 347 g 152 t
 ORIGIN

alignment_scores:
 Quality: 1106.00 Length: 309
 Ratio: 4.337 Gaps: 4
 Percent Similarity: 82.524 Percent Identity: 77.346

alignment_block:

US-09-768-877-2 x BF310408

Align seg 1/1 to: BF310408 from: 1 to: 942

216 AspGlnCysLeuIleSerCysValLeuSerProAlaGlyAlaIle 232
 3 GACCAGTGTGATCAGTGTGCTGCTGCTGACGCCACAGAGAGGTGCCG 52
 232 gGluLeuGlyIlePheHisAlaPheIleValSerAspLeuArgGluLeu 249
 53 GGAGCTGGGGAGTTCATGCTTCATGTCTCGGACTGCGGAGACTCC 102
 249 IuGlyAlaGlyIleCysIleLeuLeuArgIleGlnAspProTrp 265
 103 AGGCTCAGCGGGCCAGTGCATCTGCTGCGGATCACAACCCCTGG 152
 266 GlyArgArgCysTrpGlnGlyLeuTrpArgGluGlyGlyIleTrpSe 282
 153 GCGCGCGCTGCTGGCAGGGGCTCTGGAGAGAGGGGGGTGAAGGTGAG 202
 282 rGlnValAspAlaIleValAlaSerGluLeuSerGlnLeuGlnIle 299
 203 CCAGTACATGACAGGAGCATCTGAGCTCTCCACGTCACAGAG 252
 299 IyGluPheTrpValGluGluGluIlePheLeuArgGluPheAspGluLeu 315
 253 GGGAGTCTGGGTGGAGAGAGAGAGATTCCTCAGGAGATTGACGACTC 302
 316 ThrValGlyTrpProValThrGluAlaGlyHisLeuGlnSerLeuTrp 332
 303 ACCCTTGGCTACCCGCTACAGAGAGCGG. CACCTGACAGACCTCTAC 351
 332 rGluArgLeuLeuCysHisThrArgAlaIleuProGlyAlaTrpValIle 349
 352 AGAGAGCTGCTTGCATACGCGGGGCTGCTGCGGCTGAGGCTCAAG 401
 349 IyGlnSerAlaGlyIleCysArgAsnAsnSerGlyPheProSerAspPro 365
 402 GCCAGTCAGAGAGAGGCTGCGGAGAACAAAGAGCGCTTCCACAGACCC 451
 366 LysThrPheLeuArgValSerGluProSerGluValTrpIleAlaVal 382
 452 AAATTCGTGGCTGCGGCTCAGAACCGATGAGGTGTACATTCGCGCT 501
 382 uGlnArgSerArgLeuHisAlaIleAspTrpAlaGlyArgAlaArgAla 399

|||||
502 GCAGAGATCCAGGCTGCACGGCGG . GACTTGGCAGCGCG . GCCCGGCGCAC 549
|||||
399 euValGlYAspSerHisThrSerTrpSerProAlaSerIleProGlyLys 415
|||||
550 TGGTGSGGTGACATCATCTTCGTGGAGGCCACGCGAGCATCCCGGCGCAAG 599
|||||
416 HisTyrGlnAlaValGLyLeuHisLeuTrpLysValGIuLysArgVa 432
|||||
600 CACTACCAGGCTGTGGGTTCCTGACCTTCGAAGGTAGAGAAGCGGGG . GT 648
|||||
432 lasmLeuProArgValLeuSerMetProProValalaglyThrAlaCysH 449
|||||
649 CAATCTGCTAGGTCCTGTATGCCCGCGGGTGGACCGCTGCCATGC . 697
|||||
449 isAlaTyAspArgGIuValHisLeuArgCysGIuLeuSerProGlyTYr 465
|||||
698 ATACGAGGGAGTCACCTGCTGGGAAGCTCTTACCGGGTTAT 739
|||||
466 Tyr LeuAlaValProSerThrPheLeuLysAspAlapr 478
|||||
740 ACTGGTGCCCGACAGCTCTGGAGAGCGCAGGAGTCTCTGGCGGAGTCCTA 789
|||||
478 oGlyGIuPheLeuLeuArg 484
|||||
790 CGGGGATTCTTAGAAAGAAGGATCGAGAAACACCGGGGAGCGGCTCGGG 839
|||||
485 ValPheSerThrGlyArgValSerLeuSerAlaIleArg 497
|||||
840 TTGGGGAGCGACACGCTTCGGAACGCCCRAGGGGGA..... 874
|||||
498 AlaValalalysAsnThrThrProGly 506
|||||
875 AGAATTCGCCACACGAGGACACCGGGA 901

```

seq_name: gb_htc:AK002548

seq_documentation_block:
LOCUS      AK002548                      1137 bp      mRNA      linear      HTC 19-JAN-2002
DEFINITION Mus musculus adult male kidney cDNA, RIKEN full-length enriched
            library, clone:0610011119:caipain 10, full insert sequence.
ACCESSION  AK002548
VERSION    AK002548.1 GI:12832609
KEYWORDS   HTC; CAP trapper.
SOURCE     Mus musculus (strain:C57BL/6J) adult male kidney cDNA to mRNA,
            clone.lib:RIKEN full-length enriched mouse cDNA library
            clone:0610011119.

ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  1 (sites)
AUTHORS   Carninci,P. and Hayashizaki,Y.
TITLE     High-efficiency full-length cDNA cloning
JOURNAL   Meth. Enzymol. 303, 19-44 (1999),
MEDLINE   99279253
PUBMED    10349636

REFERENCE  2 (sites)
AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE   20493374
PUBMED    11042159

REFERENCE  3 (sites)
AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishino,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsunaga,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE     RIKEN integrated sequence analysis (RISA) system--384-format

```

sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
MEDLINE
PUBMED
11076861
4 (sites)
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001),
5 (bases 1 to 1137)
Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A.,
Arakawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bulc, C.,
Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T.,
Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hiraoka, T., Hori, F.,
Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S.,
Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K.,
Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J.,
Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,
Schraml, L., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T.,
Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,
Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yamahaka, I.,
Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and
Hayashizaki, Y.
Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,
URL: <http://genome.gsc.riken.go.jp/>,
Fax: 81-45-503-9216)
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for
further details.

```

cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
15' GAGAGAGAGCGCCCACTCGAGTGTGTTTTTTTTTTTN 3', cDNA was
prepared by using thermo-activated reverse transcriptase
and subsequently enriched for full-length by cap-trapper. cDNA went
through one round of normalization to Rot - 5.0. Second strand cDNA
was prepared with the primer adapter of sequence[5'.
GAGAGAGAGAGGATTCGAAGAGCTCAATTAATTAATAACCCCCCCCCC 3']. cDNA was
cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end:
XhoI. Host: SOLR.

Location/Qualifiers
1..1137
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="MGD:MG1:189492"
/db_xref="taxon:10090"
/clone="0610011119"
/sex="male"
/tissue_type="kidney"
/cdonc_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
99..893
/gene="Capn10"
99..893
/gene="Capn10"
/note="calpain 10
data source:MGD, source key:MG1:1344392, evidence:ISS
putative"
/codon_start=1
/protein_id="BAB22179.1"
/db_xref="GI:12832610"
/translation="NRVRAETPARELFDRFAAPSDSLFVNLSTPLAQRFEDITWR
RPEICATPOLFPDPNWEGQYKQGLLGCACALQKSOHLLDQVFPQCPGWS
QYQGFETCRWQFHWEEVITDRLPCLAGKLCFRCORDEVFWLPLEKAYKVG
SYEHLWAGADALVDLTGSAETRSKLDVTKAQGDQRPQSGGRTCRLLHLKDR
LISCVLSPRAGTANATPTSTLSLQRRKSVLERLSWAGVLLPL"

```

polya_signal 1112..1117
/note="putative"
polya_site 1137
/note="putative"
BASE COUNT 258 a 313 c 322 g 244 t
ORIGIN

alignment_scores:
Quality: 1091.00 Length: 230
Ratio: 5.005 Gaps: 0
Percent Similarity: 94.783 Percent Identity: 86.087

alignment_block:
US-09-768-877-2 x AK002548 ..

Align seg 1/1 to: AK002548 from: 1 to: 1137

```

1 MetarglaaglyargglyalathProAlaargluleuhearaspal 17
||||| 1112..1117 |||||
99 ATGGGGGGGCGGCGGAGACCGCGCGGAGCTCTCCGCGAGCGC 148
17 aalaphProAlaAlaaspSerSerleuPheCysleuSerthProL 34
||||| 1112..1117 |||||
149 GGCATTCGCCCGCGGAGCTCTCTTACACACTTGTCCAGCTTC 198
||||| 1112..1117 |||||
34 eualaglnPhearggluaspIlethTTPargargProglngluilecys 50
||||| 1112..1117 |||||
199 TGGCCGAGTTTCGGGAGACATCCTTGAGAGCAGCCCGAAGATGT 248
||||| 1112..1117 |||||
51 AlathProargleuPheProaspPhearggluileuValysgl 67
||||| 1112..1117 |||||
249 GCCACACTCAGCTGTTCCAGATACACCGAGGAGGAGGAGGAGCA 298
||||| 1112..1117 |||||
67 nglyleuLeuGLYASPCystTTPheleuCysAlaAlaLeuLnt 84
||||| 1112..1117 |||||
299 AGGCGCTGGGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 348
||||| 1112..1117 |||||
84 ysearTghIstleuLeuaspPglValIleProglnglnProSerTTP 100
||||| 1112..1117 |||||
349 AGATCACAACCTCTGAGCAGGCTCTCCCTCAGAGCAGCCAGGCT 398
||||| 1112..1117 |||||
101 AlaspPglngluthTarglySerPheThCysargIlethPglngl 117
||||| 1112..1117 |||||
399 TCGACACAGAAATACCAAGGCTCTTCCACTGTCGAGATTGGCAG 448
||||| 1112..1117 |||||
117 yATGTTPValgluValThrThraspaspargleuProCysleuAla 134
||||| 1112..1117 |||||
449 ACACCTGGAGAGATGACCATATGATGCTGCTGCTGCTGCTGCTG 498
||||| 1112..1117 |||||
134 rgluLeuCysPheSerArgCysGlnarggluaspValPheTTPleu 150
||||| 1112..1117 |||||
499 GACTCTGCTTTTCCCGGTCGACAGAGAGATGTTGGCGCTTCCCT 548
||||| 1112..1117 |||||
151 leuGlulysValTyrAlaIysValIhSgIysertTyrGlnIstleu 167
||||| 1112..1117 |||||
549 CTGGAAAGAGCCATATGCTAAGGCTCATGATGATGATGATGATG 598
||||| 1112..1117 |||||
167 aclyGlnValAlaaspAlaLeuValaspLeuPheThrglyleuAla 184
||||| 1112..1117 |||||
599 AGGCGAGAGGAGATGCTTACTGATGCTCAGTGGAGCCGCGAGAA 648
||||| 1112..1117 |||||
184 rGTTPasnleuLysglYValAlaaglySerGlyglnglnInasparg 200
||||| 1112..1117 |||||
649 GCTGGAGCTTGAAGATGTACGAAAGCCAGGCGGAGAGAGAGACC 698
||||| 1112..1117 |||||
201 GYAAGTTPGlnIhSargThrCysargGlnleuLeuIstleuLysasp 217
||||| 1112..1117 |||||
699 ACTGTGGGAGGACAGAACTTGTGCGCAGCTACCTCAGTAAGAGCG 748
||||| 1112..1117 |||||
217 nCysleuIstleuSerCysValleuSerProargIlaagly 230
||||| 1112..1117 |||||
749 GTGTCTAAATCAGCTCTGTGTCTTACCGCCAGAGCAGAGT 788
||||| 1112..1117 |||||

```

seq_name: gb_hlc:AK013497

seq_documentation_block: 1269 bp mRNA linear HTC 19-JAN-2002
AK013497
LOCUS
DEFINITION Mus musculus adult male hippocampus cDNA, RIKEN full-length
enriched library, clone:2900008C09:calpain 10, full insert
sequence.

ACCESSION AK013497
VERSION AK013497.1 GI:12850882
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male hippocampus cDNA to mRNA,
clone.lib:RIKEN full-length enriched mouse cDNA library
clone:2900008C09.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (sites)
REFERENCE
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636

REFERENCE
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159

REFERENCE
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Kono,H., Akiyama,J., Nishi,K., Kitsuana,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861

REFERENCE
AUTHORS 4 (sites)
TITLE The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
JOURNAL Functional annotation of a full-length mouse cDNA collection
NATURE 409, 685-690 (2001)
MEDLINE 11076861
PUBMED 11076861

REFERENCE
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
Aikawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanganu,T.,
Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,
Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,
Kato,H., Kawai,J., Kojima,Y., Kono,H., Kouda,M., Koya,S.,
Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
Schrabel,L., Shibata,K., Shibata,Y., Shingawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamana,K., I.,
Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
Hayashizaki,Y.

TITLE

JOURNAL

Submitted (10-JUL-2000) Yoshihide Hayashizaki. The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
Kanagawa 230-0045, Japan (E-mail:genome-res@gsr.riken.go.jp,
URL: <http://genome.gsc.riken.go.jp/>, Tel:81-45-503-9222,
Fax:81-45-503-9216)
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for

COMMENT

1 MetArgAlaGlyArgGlyAlaThrProAlaAraGluLeuPheArgAspAl 17
110 ATGGGGCGCTGGCCGGAGACGCCGGCGGGAGCTTTCGGGACGC 159
17 aaLaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThr.ProL 34
160 GGCATTCCC GCCACTCTCGCTCTCTTTTACACCTGTCTCACGCCCTC 209
34 euLaGlnPheArgGluAspIleThrTrpArgArgProGlnGluLeuCys 50
210 TGCCCCAGTTTCGGGAGGACATCATCTTGAGACGACC CCAGGAATACTGT 259
51 AlaThrProArgLeuPheProAspAspProArgGluGluGlnVallysGl 67
260 GCCACACCTCAGCTGTTTCCAGATAACCCATGCGGGAGGACAGGTGAAGCA 309
67 nGlyLeuLeuGlyAspCystTrpPheLeuCysAlaCysAlaLeuGlnL 84
310 AGGCGTCTCGGAGATTGCTGTCTCTGTGTGCTCTGGCGCTTCAGA 359

10/10/10

10/10/10

```
/clone="IMAGE:5118806"
```

/tissue_type="pooled lung tumors"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: lung; Vector: pTV3D-Pac (Pharmacia) with a
 modified polylinker; Site_1: NotI; Site_2: EcoRI; 1st
 strand cDNA was prepared from mRNA obtained from pooled
 lung tumors with a Not I - oligo(dT) primer [5',
 TGTACCAATCTGAGAGGAGCGCCCTCTTTTCTTTTCTTTT 3'].
 Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not
 I and Eco RI sites of the modified pTV3 vector. Library
 went through one round of normalization, and was
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 216 a 289 c 304 g 219 t

alignment_scores:

Quality: 1078.00 Length: 231
 Ratio: 4.945 Gaps: 1
 Percent Similarity: 94.372 Percent Identity: 85.714

alignment_block:

US-09-768-877-2 x B1410486 ..

Align seg 1/1 to: B1410486 from: 1 to: 1028

```

1 MetATGAlaGlyAArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
  |||||
93 ATGCGGGCGGTCGGGCGGAGAGCGGCGGCGGAGCTTCGCGGAGCC 142
17 AlaAheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrPro 34
  |||||
143 GGCATTCCCGCGGCTGCTCTCTTTACAACTTGTCCAGCCCTC 192
  |||||
34 euaAglnPheArgGluAspIleThrPArgArgProGlnIleCys 50
  |||||
193 TGGCCAGTTTCGGAGAGACATCACTGGAGACGCCAGAAATCTGT 242
  |||||
51 AlaThrProArgLeuPheProAspAspProArgGluGlnValLysGln 67
  |||||
243 GCCACACTTCAGCTGTTCCAGATACCCATGCGGAGGAGGAGTGAACA 292
  |||||
67 nGlyLeuLeuGlyAspCysTrpPheLeuCysAlaCysAlaIleuGlnL 84
  |||||
293 AGGGCTGCTGGAGATTCTGCTCTCTGTGCTGCTGCTGCTGCTGCTG 342
  |||||
84 ySserAghHisLeuLeuAspGlnValIleProProGlnLysProSerTrp 100
  |||||
343 AGACTCAACACCTCTCGACCAAGCTCTCTCCAGGAGACAGGCGTGG 392
  |||||
101 AlaAspGlnGluTrpArgGlySerPheThrCysArgGlyIleTrpGlnPheGln 117
  |||||
393 TCTACCAAGAAATACCAAGGCTTCTCACTGTGAGATTGGAGATTGG 442
  |||||
117 yArgTrpValGlnValThrThrAspAspArgLeuProCysLeuAlaGly 134
  |||||
443 ACACTGGAGAGAGTGAACATAGATGATCGTCTCTCTCTCTCTCTCT 492
  |||||
134 rGlyLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
  |||||
493 GACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 542
  |||||
151 LeuGlnLysValTyrAlaLysValHisGlySerTrpGlnHisLeuTrpAl 167
  |||||
543 CTGGAAGAAAGGCTTATGCTTAAGGTCATGATCGTATGAGACCTGTGG 592
  |||||
167 aGlyGlnValAlaAspAlaLeuValaAspLeuThrGlyLysLeuAlaGln 184
  |||||
593 AGGGCAAGTGGCAGATGCTTAGTGAATCTCAGTGAAGCTGGACAGAAA 642
  |||||
184 rGTrpAsnLeuLysGlyValAlaGlySerGlyGlnGlnAspArgPro 200
  |||||
643 GGTGGAGCTTGAAGATGTAACAAAGCCAGCGGCCAGGAGAGAGAGCC 692
  |||||

```

201 GLyArgTrpGlnHisArgThrCysArgGlnLeuLeuHisLeuLysAsp 217
 :|||
 693 AGGTCGGGAGGAGCAGACACTGTGCGACACTCTCCACTGAGAGAGCC 742

217 IncSLeuLeuSerCysCysValLeuSerProArgAlaGly 230
 :|||
 743 GGTGTCTATTCAGCTCTCTGCTCTTACGCCAGACAGAGCT 783

seq_name: gb_est2:BE390603

seq_documentation_block:

LOCUS BE390603 601 bp mRNA linear EST 21-JUL-2000
 DEFINITION 601284961F1 NIH_MGC_44 Homo sapiens cDNA clone IMAGE:360638 5',
 mRNA sequence.
 ACCESSION BE390603
 VERSION BE390603.1 GI:9335968
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 601)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: c9aaps-remail.nih.gov
 Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 Plate: LNCM256 row: 3 column: 15
 High quality sequence stop: 600.

FEATURES

source

1..601
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NIH_MGC_44"
 /tissue_type="endometrium, adenocarcinoma cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: uterus; Vector: pOTB7; Site_1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(C). Library constructed by Ling Hong
 in the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 104 a 173 c 225 g 99 t

alignment_scores:

Quality: 1076.00 Length: 197
 Ratio: 5.462 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-768-877-2 x BE390603 ..

Align seg 1/1 to: BE390603 from: 1 to: 601

```

200 ProGlyArgTrpGlnHisArgThrCysArgGlnLeuLeuHisLeuLysAs 216
  |||||
3 CCAGGCCGCTGGGAGCAGACACTTGTGCGCAGTGTCTCCACCTGAAGGA 52
  |||||
216 pGlnCysLeuLysSerCysValLeuSerProArgAlaGlyValaArg 233
  |||||
53 CCAGTGTCTGATCAGCTGTGCTGCTCAGCCAGAGAGAGGTGCCGGG 102
  |||||

```

233 luLeuGlyGluPheHisAlaPheIleValSerAspLeuArgGluLeuGln 249
103 AGCTGGGGCAGTTCCATGCTTCGGACCTCGGGAGCTCCAG 152
250 GlyAlaGlyGlnCysIleLeuLeuLeuArgIleGlnAsnProTrpG1 266
153 GGTACGGCGGGCAGTGCATCCTGCTGCGGATCCAGAACCCCTGGGG 202
266 yArgArgCysTrpGlnGlyLeuTrpArgGluGlyGlyGluGlyTrpSerG 283
203 CCGCGGTGCTGCGAGGGGCTCTGGAGAGGGGGGTGAAGGTGGAGCC 252
283 InValAspAlaAlaValAlaSerGluLeuLeuSerGlnLeuGlnGly 299
253 AGGTAGATCGACGGTAGCATCTGAGTCTCTGCCAGCTCCAGGAAGG 302
300 GluPheTrpValGluGluGluGluPheLeuArgGluPheAspGluLeuTh 316
303 GAGTCTGGTGGAGAGGAGGAGTCTCTCAGGGAGTTGACAGAGCTCAC 352
316 rValGlyTrpProValThrGluAlaGlyHisLeuGlnSerLeuTrpThrG 333
353 CGTTGGCTACCGGTCACGGAGGGCGGCCACCTGCAGAGGCTCTACACAG 402
333 luArgLeuLeuCysHisThrArgAlaLeuProGlyAlaTrpValLysGly 349
403 AGAGGCTGCTCTGCCATACGCGGGCGCTGCCTGGGGCTGGGTCAAGGC 452
350 GlnSerAlaGlyGlyCysArgAsnAsnSerGlyPheProSerAsnProLy 366
453 CAGTCAGCAGGAGGCTGCCGGAACACAGACGGGTTCTCCAGCAACCCAA 502
366 sPheTrpLeuArgValSerGluProSerGluValTyrIleAlaValLeu 383
503 ATTCTGGCTGGGGTCTCAGAACCGAGTGAAGTGTACATTGGCTCTGCTGC 552
383 luArgSerArgLeuHisAlaAlaAspTrpAlaGlyArgAla 396
553 AGAGATCCAGGCTGCACGCGCGGACTGGGCGAGCGCGGGCC 593

name: qb_est2:B1412050

seq_documentation_block:	
LOCUS	BI412050 940 bp mRNA linear EST 14-AUG-2001
DEFINITION	G0265960F1 NCI_CGAP_Lu33 Mus musculus cdna clone IMAGE:5121446 5', mRNA sequence.
ACCESSION	BI412050
VERSION	BI412050.1 GI:15172973
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	house mouse
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;
AUTHORS	1 (bases 1 to 940)
TITLE	NIH-MGC http://mgc.nci.nih.gov/ .
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT	Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: Gilbert Smith, Ph.D. CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM1296 row: g column: 15 High quality sequence start: 3 High quality sequence stop: 858. Location/Qualifiers 1..940
FEATURES	
source	

```

/organism="Mus musculus"
/strain="CZECH II"
/db_xref="taxon:10090"
/clone="IMAGE:5121446"
/clone_lib="NCI_CGAP_Lu33"
/tissue_type="pooled lung tumors"
/lab_host="DHL0B (phage-resistant)"
/notes="Organ: lung; Vector: pT730-Pac (Pharmacia) with a
modified polylinker; Site.1: NotI; Site.2: EcoRI; 1st
strand cDNA was prepared from mRNA obtained from pooled
lung tumors with a NotI - oligo(dT) primer [5'
TCGTACCAACTGTAAGTGGAGCGCCGCTCTCTTTTTTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo. "
193 a 269 c 275 g 203 t

```

```

alignment_scores:
  Quality: 1068.50      Length: 289
  Ratio: 4.586          Gaps: 4
  Percent Similarity: 80.623  Percent Identity: 70.934

alignment_block:
  US-09-768-877-2 x B1412050  ..

Align seg 1/1 to: B1412050 from: 1 to: 940

1 MetArgAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
58 ATGCGGGCGGTCCGGCCGAGAGCGCGCGGGAGCTCTTCGGGAGCG 107
17 aalapeProAlaAlaAspSerLeuPheCysAspLeuSerThrProL 34
108 GCATATCCCGGCTCGGACTCCCTGCGCTCTTTACAACTGTGCCAGCGCTC 157
34 euAlaGlnPheArgGluAspIleThrTrpArgProGlnGluIleCys 50
158 TGGCCCGAGTTTCGGGAGGACATCACATTGGAGACACCCAGCAAAATCTGT 207
51 AlaThrProArgLeuPheProAspAspProArgGluGlyGlnValLysG 67
208 GCCACACCTTCAGCTGTTTCAGATAACCCATGGGAGGACAGGTGAAGCA 257
67 nGlyLeuLeuGlyAspCysTrpPheLeuCysAlaCysAlaAlaLeuGlnL 84
258 AGGGCTGCTGGGAGATGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 307
84 ysSerArgHisLeuLeuAspGlnValIleProProGlyGlnProSerTrp 100
308 AGAGTCAACACCTCTGGACCAGGTCTTCCTCCAGACACAGCGCGTGG 357
101 AlaAspGlnGluTyrArgGlySerPheThrCysArgIleTrpGlnPheG 117
358 TCTGACCAAAATACCAAGGCTTCTTCACCTGTTCGGATTTGGCAGTTTGG 407
117 yArgTrpValGluValThrThrAspAspArgLeuProCysLeuAlaGlyA 134
408 ACACTGGGAGGAAGTGACCATAGATATCGTCTGCTGCTGCTGCTGCTGCTG 457
134 rGLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
458 GACTCTGCTTTTCCGGTCCGAGAGAGAGATGTGTTCTGCTGCTTCCCTTA 507
151 LeuGluLysValTyrAlaLysValHisGlySerTyrGluHisLeuTrpAl 167
508 CTGGAAGAGGCTATGCTAAGGTCCATGGATCGTATCAGCACCTGTGGGC 557
167 aGlyGlnValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGluA 184

```

```

558 AGGCAAGTGGCAGATGCTTAGTGAGATCTCACTGCAAGCTGCGAGAAA 607
184 TGTPrAsnLeuLysGly.ValAlaGlySerGlyGlyGlnGlnAspArgPr 200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
608 GGTGACGCTTCAGAGATGTACGAAAGCAGCCGCGCAGACAGACAGACC 657
200 OGlyArgTPrGlnHisArgThrCysArgGlnLeuLeuHisLeuLysAsp 217
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
658 CAGTGTGGGAGACAGAACTTGTGGCAGCTACTCCACTGGAAGAGACC 707
217 InCysLeuLeuSerCysCysValLeuSer.ProArgAlaGly..... 230
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
708 GGTGTCTATACAGCTGCTGTGTGTAGCCCGCCAGAGCAGGTACAGCTAT 757
231 .....AlaArgLeuLeuGly.. 235
758 GGCTACACTCCCTCCACACATTAGCTGTATACAGACGAAAGAAATCGGTC 807
235 ..... 235
808 CTTGAGAGGCTGTCTGTGGTGGGTGTCTGTCTGTACCCCTCCTAG 857
236 .....GlnPheHisAlaPheIleVal 242
858 AGAGAACTATACACTTCTCTCCTAGACTTATCTTTCACAGCTAGGCGCA 907
242 lSerAspLeuArg 246
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
908 AGCGAGCTACGT 920
seq_name: gb_est2:BE619302

seq_documentation_block:
LOCUS BE619302 789 bp mRNA linear EST 20-OCT-2000
DEFINITION 601473184F1 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:3876141 5',
mRNA sequence.
ACCESSION BE619302
VERSION BE619302.1 GI:9890240
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 789)
NIH-MGC http://mhc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DTF/Gazdar
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LAM9636 row: g column: 22
High quality sequence stop: 663.
Location/Qualifiers
1..789
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3876141"
/clone_lib="NIH_MGC_68"
/tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."
BASE COUNT 145 a 218 c 295 g 131 t
ORIGIN

```

```

alignment_scores:
  Quality: 1056.50      Length: 251
  Ratio: 4.439          Gaps: 7
  Percent Similarity: 94.821  Percent Identity: 92.430

alignment_block:
US-09-768-877-2 x BE619302
Align seg 1/1 to: BE619302 from: 1 to: 789

36 GlnPheArgGluAspGlnValIleProProGlnProGlnGlnIleCysAlaLeu 52
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
13 CAGTCCCGCAGAGACATACCTGGAGGCGGCCCGCCAGAGAT.TGTGCCAC 61
52 rProArgLeuPheProAspAspProArgGlnGlnValIleLysGlnGlyL 69
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
62 ACCCGGCTGTTCAGATGACCCAGCGGGAAGGCGAGGTGAAGCAGGCGC 111
69 euleuGlyAspCysTPrPheLeuLysCysAlaCysAlaAlaLeuGlnLysSer 85
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
112 TGTGGGGGA.TGTGGTTCCTGTGTGTCTGCTGCGCGCTCCAGAAAGAGC 160
86 ArgHisLeuLeuAspGlnValIleProProGlnProGlnProGlnAlaAs 102
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
161 AGGACCTCCCTCGACACAGATCATCTCCGGGACAGCGGAGCTGGGCGCA 210
102 pGlnGlnTyrArgGlySerPheThrCysArgIleTPrGlnPheGlyArgT 119
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
211 CCAGGAGTACCGGGGCTCTTCACCTGTGGCATTTGGAGAT.GGAGGCT 259
119 rPValGlnValThrThrAspAspArgLeuProCysLeuAlaGlyArgLeu 135
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
260 GGGTGGAGGTACACAGATATACCGCTGCGCTGCTGCTGCGGAGAGACTC 309
136 CysPheSerArgCysGlnArgGluAspValPheThrLeuProLeuLeuG 152
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
310 TGTTCCTCCCGCTCCAGAGGAGGAGATGTGTTCCTCCCTTACT.GA 358
152 uLysValTyrAlaLysValHisGlySerTyrGlnHisLeuTPrAlaGly 169
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
359 AAAGCTACGCCCAAGCTCCATGGGTCTACAGACACTGTGGGCGGCGC 408
169 InValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGlyArgT 185
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
409 AGGTGGCGAGATCCCTGTGTGACTGACCGGCGCTG.GCAGAAAGATGG 457
186 AsnLeuLysGlyValAlaGlySerGlyGlnGlnAspArgProGlyArg 202
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
458 AACCTGAAGGCGCTACAGAGAGCGGAGGCCAGAGACAGGCGGCGG 507
202 gTPrGlnHisArgThrCysArgGlnLeuLeuHisLeuLysAspGlnCys 218
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
508 CTGGAG.CACAGACTGTGCGGAGCTGTCCACTCAAGAGACAGAGGT 556
219 LeuLeuSerCysCysValLeuSerProArgAlaGlyAlaArgGlnLeuG 235
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
557 CTGATCAGCTGCTGCGCTGCTGACCGCCAGAGAGAGGTGCCCGGAGCTGG 606
235 LysGlnPheHisAlaPheIleValSerAspLeuArgGlnLeuGlnGly 251
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
607 GGAAGTTCCTATCTTCATGTCTCGGAGCTGGGGAAGCTCCAGAGTCA 656
251 nAlaGlyGlnCysIleLeuLeuLeuArgIleGlnAsnProTPrGlyArg 267
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
657 GGGGGGCAATGCATC...CTGCTGTGGATCAAGAAACCCG...GGGCG 700
268 ArgCysTPrGlnGlnLeuTPrArgGlnGlyGlnGlyLysTPrSer 282
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
701 GGGGTGTGGCAGGCGGCTG...GAAAGGGGGTGAAGAGGTGGGCG 742
seq_name: gb_est1:AL558905

```



```

seq_documentation_block:
LOCUS       AL558905                               776 bp    mRNA    linear    EST 16-FEB-2001
DEFINITION  AL558905 LTI_NFL008_Tc2 Homo sapiens cDNA clone CS0DJ007YG13 5
            prime, mRNA sequence.
ACCESSION   AL558905
VERSION     AL558905.1 GI:12903884
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 776)
AUTHORS    Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
TITLE      Full-length cDNA libraries and normalization
JOURNAL    Unpublished (2001)
COMMENT    Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1..776
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DJ007YG13"
                     /clone_lib="LTI_NFL008_Tc2"
                     /sex="male"
                     /tissue_type="T cells from T cell leukemia"
                     /note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector. Library was normalized. Library was constructed by
                     Life Technologies. Contact : Feng Liang Life Technologies,
                     a division of Invitrogen 9800 Medical Center Drive
                     Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
                     Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT      119 a   237 c   292 g   124 t
ORIGIN

```

alignment_scores:

```

Quality: 1047.00      Length: 196
Ratio: 5.425          Gaps: 0
Percent Similarity: 98.469 Percent Identity: 97.959

```

alignment_block:

```
US-09-768-877-2 x AL558905
```

```
Align seg 1/1 to: AL558905 from: 1 to: 776
```

```

1 MetArgAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
|||||
187 ATCGGGCGGGCGGGCGGCGGAGCGGCGGAGGAGCTGTTCCGGGACGC 236

17 alaIaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrProL 34
|||||
237 CCCTCTCCCGCGCGGAGCTCCTCGCTCTTCTGCGACTTGCTACGCGCG 286

34 euAlaGlnPheArgGluAspIleThrTrpArgArgProGlnGluIleCys 50
|||||
287 TGGCCAGCTTCCCGGAGGACATCAGTGGAGCGGCGGCGGAGGATTTGT 336

51 AlaThrProArgLeuPheProAspAspProArgGluGlyGlnValLysG1 67
|||||
337 GCCACACCCCGGCTGTTTCCAGATGACCCACGGGAAGGCGAGGTGAACCA 386

67 nGlyLeuLeuGlyAspCysTrpPheLeuCysAlaCysAlaAlaLeuGlnL 84
|||||
387 GGGGCTGTGGGGGATGCTGGTCTCTGCTGCTTCGCTTCGCCACGCTGCAGA 436

84 ySeraArgHisLeuLeuAspGlnValIleProProGlyGlnProSerTrp 100
|||||

```

```

437 AGAGCAGGACACCTCTCTGGACCAGGTTCATTCTCCGGGACAGCCGAGCTGG 486
101 AlaAspGlnGluTyrArgGlySerPheThrCysArgIleTrpGlnPheG1 117
|||||
487 GCCGACCAGGAGTACCGGGGCTCCTTACCTGTCGCAATTGGCAGTTGG 536

117 yArgTrpValGluValThrThrAspAspArgLeuProCysLeuAlaGlyA 134
|||||
537 AGCTGGGTGGAGGTGACCCACAGATGACCGCTGCCGTGCCCTTGCAGGGA 586

134 rGluCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
|||||
587 GACTCTGTTTCTCCCGCTGCCAGAGGAGGAGATGTTCTTGGCTCCCTTA 636

151 LeuGluLysValTyrAlaLysValHisGlySerTyrGluHisLeuTrpAl 167
|||||
637 CTGGAAAGGTCTACGCCAAGGTCCATGGGTCTCTACGAGCACTGTGKGC 686

167 aGlyGlnValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGluA 184
|||||
687 CGGGCAGGTGGCGGATGCCCTGTTGACCTGACCGGCGGCTGGCAGAAA 736

184 rGTrpAsnLeuLysGlyValAlaGlySerGlyGlyGln 196
|||||
737 GATGAACCTGAARGCGGTACAGGGAAGCGGAGGCCAG 774

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2002, 10:34:46 ; Search time 14.65 Seconds
(without alignments)
1120.409 Million cell updates/sec

Title: US-09-768-877-2

Perfect score: 3649
Sequence: 1 MRAGCATPARELFRDAFP.....SOENIGOFLOEVSVAWVKT 672

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/pdata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/pdata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/pdata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/pdata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/pdata/2/1aa/PCRTUS.COMB.pep:*
6: /cgn2_6/pdata/2/1aa/Dackfilest.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3649	100.0	672	US-09-422-869-2	Sequence 2, Appl1
2	2963	81.2	666	US-09-422-869-18	Sequence 18, Appl1
3	2715.5	74.4	517	US-09-422-869-6	Sequence 6, Appl1
4	2701.5	74.0	544	US-09-422-869-4	Sequence 4, Appl1
5	2699	74.0	513	US-09-422-869-8	Sequence 8, Appl1
6	2336	64.0	444	US-09-422-869-10	Sequence 10, Appl1
7	858.5	23.5	274	US-09-422-869-12	Sequence 12, Appl1
8	635	17.4	703	US-08-835-099A-1	Sequence 1, Appl1
9	635	17.4	703	US-09-157-349-1	Sequence 2, Appl1
10	635	17.4	712	US-08-835-099A-2	Sequence 1, Appl1
11	635	17.4	712	US-09-157-349-2	Sequence 2, Appl1
12	618.5	16.9	821	US-09-422-869-22	Sequence 22, Appl1
13	618	16.9	703	US-09-422-869-24	Sequence 24, Appl1
14	612	16.8	703	US-09-422-869-27	Sequence 27, Appl1
15	604	16.6	700	US-08-726-525-7	Sequence 7, Appl1
16	604	16.6	700	US-08-487-942-7	Sequence 7, Appl1
17	604	16.6	700	US-08-487-942-7	Sequence 7, Appl1
18	604	16.6	700	US-08-726-036A-7	Sequence 7, Appl1
19	604	16.6	700	US-09-422-869-23	Sequence 23, Appl1
20	584	16.0	639	US-09-083-516-7	Sequence 7, Appl1
21	561.5	15.4	690	US-09-422-869-25	Sequence 25, Appl1
22	513.5	14.1	139	US-09-422-869-14	Sequence 14, Appl1
23	483.5	13.3	641	US-09-422-869-26	Sequence 26, Appl1
24	472	12.9	138	US-09-422-869-16	Sequence 16, Appl1
25	271	7.4	854	US-08-928-692-17	Sequence 17, Appl1
26	264.5	7.2	842	US-08-928-692-18	Sequence 18, Appl1
27	133	3.6	67	US-09-120-365-97	Sequence 97, Appl1

28	133	3.6	67	4	US-09-515-039-97	Sequence 97, Appl1
29	124	3.4	67	3	US-09-120-365-95	Sequence 95, Appl1
30	124	3.4	67	4	US-09-515-039-95	Sequence 95, Appl1
31	121	3.3	67	3	US-09-120-365-98	Sequence 98, Appl1
32	121	3.3	67	4	US-09-515-039-98	Sequence 98, Appl1
33	117.5	3.2	67	3	US-09-120-365-99	Sequence 99, Appl1
34	117.5	3.2	67	4	US-09-515-039-99	Sequence 99, Appl1
35	117	3.2	67	3	US-09-120-365-96	Sequence 96, Appl1
36	117	3.2	67	4	US-09-515-039-96	Sequence 96, Appl1
37	115	3.2	67	3	US-09-120-365-94	Sequence 94, Appl1
38	115	3.2	67	4	US-09-515-039-94	Sequence 94, Appl1
39	110	3.0	3519	4	US-08-804-227C-9	Sequence 4, Appl1
40	105	2.9	1996	2	US-08-804-198-3	Sequence 9, Appl1
41	105	2.9	1996	2	US-08-804-198-3	Sequence 9, Appl1
42	102.5	2.8	512	4	US-09-356-818A-2	Sequence 2, Appl1
43	102.5	2.8	548	1	US-08-247-902A-2	Sequence 2, Appl1
44	102.5	2.8	687	5	PCT-US91-09784-2	Sequence 2, Appl1
45	100.5	2.8	505	1	US-08-220-603A-10	Sequence 10, Appl1

ALIGNMENTS

```
RESULT 1
US-09-422-869-2
; Sequence 2, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: COX, NANCY J.
; APPLICANT: SRENNAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: OTANI, KENICHI
; APPLICANT: HANIS, CRAIG L.
; APPLICANT: BELL, GRAEME I.
; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
; FILE REFERENCE: ARCD:307
; CURRENT APPLICATION NUMBER: US/09/422,869
; CURRENT FILING DATE: 1999-10-21
; EARLIER APPLICATION NUMBER: 60/134,175
; EARLIER FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 672
; TYPE: PRT
; ORGANISM: Human
US-09-422-869-2

Query Match      100.0%: Score 3649; DB 4; Length 672;
Best Local Similarity 100.0%: Pred. No. 0;
Matches 672; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY	1	MRAGCATPARELFRDAFPADSSIFCDLSTPLAQFREDITWRROECATPRLEPPDP	60
DB	1	MRAGCATPARELFRDAFPADSSIFCDLSTPLAQFREDITWRROECATPRLEPPDP	60
QY	61	REGVQKQGLGSCWELCAALQKSHLLDQYIPGQPSMAQOEYRGSEFCRIMQGRWY	120
DB	61	REGVQKQGLGSCWELCAALQKSHLLDQYIPGQPSMAQOEYRGSEFCRIMQGRWY	120
QY	121	EYTDRLPCLAGRLCFSCRCOREDFWMLPLEKKYAVAGSYEHLMAGQVADALVDLGG	180
DB	121	EYTDRLPCLAGRLCFSCRCOREDFWMLPLEKKYAVAGSYEHLMAGQVADALVDLGG	180
QY	181	IAERNLKGAVAGSGGQDPGRMEHRTCRQLHLMDCLISCCVSPRAGARELGEFFAF	240
DB	181	IAERNLKGAVAGSGGQDPGRMEHRTCRQLHLMDCLISCCVSPRAGARELGEFFAF	240
QY	241	IYSDRLPCLAGRLCFSCRCOREDFWMLPLEKKYAVAGSYEHLMAGQVADALVDLGG	300

Db 241 IVSDRELQAGQCILLRLIQRNPMGRRCWQGLWREGGEGNSQVDNAVASLLSQLEGE 300
Qy 301 FWVEEELREFDELTVGYPVTEAGHLQSLYTERLLCHTRALPGAWKQSGAGGCRNNSG 360
Db 301 FWVEEELREFDELTVGYPVTEAGHLQSLYTERLLCHTRALPGAWKQSGAGGCRNNSG 360
Qy 361 FPSNPKFWLRVSEPEVYIIVLQSRRLHAADWAGARALVGDSTSWSPASIPGKHQYAV 420
Db 361 FPSNPKFWLRVSEPEVYIIVLQSRRLHAADWAGARALVGDSTSWSPASIPGKHQYAV 420
Qy 421 GLHLKVKERRRNVLPVLSMPVAGTACHAYDREHVLRCLESPGYIYLVPSFLKADAGE 480
Db 421 GLHLKVKERRRNVLPVLSMPVAGTACHAYDREHVLRCLESPGYIYLVPSFLKADAGE 480
Qy 481 FLLRVFSTGRVSLSAIRAVAKNTTPGAALPAGEWGTVQLRGSWRVGOTAGGSRNFASYPT 540
Db 481 FLLRVFSTGRVSLSAIRAVAKNTTPGAALPAGEWGTVQLRGSWRVGOTAGGSRNFASYPT 540
Qy 541 NCFPFSPVEGPGPCRVITLHOHCRPSDTEPHPIGFHIFQVPEGGRSQDAPPLLOEPL 600
Db 541 NCFPFSPVEGPGPCRVITLHOHCRPSDTEPHPIGFHIFQVPEGGRSQDAPPLLOEPL 600
Qy 601 LSCVPHRYAQEVSRLCLLPAGTYKVPSTYLPDTEGAFVTIATRIDRPSIHSQEMLGOF 660
Db 601 LSCVPHRYAQEVSRLCLLPAGTYKVPSTYLPDTEGAFVTIATRIDRPSIHSQEMLGOF 660
Qy 661 LOEVSVMVMYK 672
Db 661 LOEVSVMVMYK 672

RESULT 2
US-09-422-869-18
; Sequence 18, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: ODA, NAOHISA
; APPLICANT: COX, NANCY J.
; APPLICANT: SREENAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: OTANI, KENICHI
; APPLICANT: HANIS, CRAIG L.
; APPLICANT: BELL, GRAEME I.
; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
; FILE REFERENCE: ARCD:307
; CURRENT APPLICATION NUMBER: US/09/422,869
; EARLIER FILING DATE: 1999-10-21
; EARLIER FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 666
; TYPE: PRT
; ORGANISM: Mus musculus.
US-09-422-869-18

Query Match 81.2%; Score 2963; DB 4; Length 666;
Best Local Similarity 81.1%; Pred. No. 3.7e-289;
Matches 544; Conservative 45; Mismatches 76; Indels 6; Gaps 2;
Qy 1 MRAGRGATPARELFDAAPFADSSLSFCDLSTPLAQFREDITWRPQICATPRLFPDPP 60
Db 1 MRVRAETPARELFDAAPFADSSLSFYNLSTPLAQFREDITWRPQICATPQLFPDPP 60
Qy 61 REGQVKQGLLDCWFLCACLQKSRHLLDQVIPPQGSWADQYRGSTCRIMQFGRW 120
Db 61 WEGQVKQGLLDCWFLCACLQKSRHLLDQVIPPQGSWADQYRGSTCRIMQFGRW 120

Qy 121 EVTTDDRLPCLAGRLCFRCQREDVFWLPLEKVKYAKVHGSYEHLMWAGQVADALVDLTGG 180
Db 121 EVTTDDRLPCLAGRLCFRCQREDVFWLPLEKVKYAKVHGSYEHLMWAGQVADALVDLTGS 180
Qy 181 LAERNLKVAGSGGQDRPGRWEHRTCRQLHLKQDCLISCCVLSPRAGARELGEFHAF 240
Db 181 LAERNLKVAGSGGQDRPGRWEHRTCRQLHLKQDCLISCCVLSPRAGARELGEFHAF 240
Qy 241 IVSDRELQAGQCILLRLIQRNPMGRRCWQGLWREGGEGNSQVDNAVASLLSQLEGE 300
Db 241 IVSDRELQAGQCILLRLIQRNPMGRRCWQGLWREGGEGNSQVDNAVASLLSQLEGE 300
Qy 301 FWVEEELREFDELTVGYPVTEAGHLQSLYTERLLCHTRALPGAWKQSGAGGCRNNSG 360
Db 301 FWVEEELREFDELTVGYPVTEAGHLQSLYTERLLCHTRALPGAWKQSGAGGCRNNSG 360
Qy 361 FPSNPKFWLRVSEPEVYIIVLQSRRLHAADWAGARALVGDSTSWSPASIPGKHQYAV 420
Db 361 FPSNPKFWLRVSEPEVYIIVLQSRRLHAADWAGARALVGDSTSWSPASIPGKHQYAV 420
Qy 421 GLHLKVKERRRNVLPVLSMPVAGTACHAYDREHVLRCLESPGYIYLVPSFLKADAGE 480
Db 421 GLHLKVKERRRNVLPVLSMPVAGTACHAYDREHVLRCLESPGYIYLVPSFLKADAGE 480
Qy 481 FLLRVFSTGRVSLSAIRAVAKNTTPGAALPAGEWGTVQLRGSWRVGOTAGGSRNFASYPT 540
Db 481 FLLRVFSTGRVSLSAIRAVAKNTTPGAALPAGEWGTVQLRGSWRVGOTAGGSRNFASYPT 540
Qy 541 NCFPFSPVEGPGPCRVITLHOHCRPSDTEPHPIGFHIFQVPEGGRSQDAPPLLOEPL 600
Db 541 NCFPFSPVEGPGPCRVITLHOHCRPSDTEPHPIGFHIFQVPEGGRSQDAPPLLOEPL 600
Qy 601 LSCVPHRYAQEVSRLCLLPAGTYKVPSTYLPDTEGAFVTIATRIDRPSIHSQEMLGOF 660
Db 601 LSCVPHRYAQEVSRLCLLPAGTYKVPSTYLPDTEGAFVTIATRIDRPSIHSQEMLGOF 660
Qy 661 LOEVSVMVMYK 671
Db 661 LOEVSVMVMYK 665

RESULT 3
US-09-422-869-6
; Sequence 6, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: ODA, NAOHISA
; APPLICANT: COX, NANCY J.
; APPLICANT: SREENAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: OTANI, KENICHI
; APPLICANT: HANIS, CRAIG L.
; APPLICANT: BELL, GRAEME I.
; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
; FILE REFERENCE: ARCD:307
; CURRENT APPLICATION NUMBER: US/09/422,869
; EARLIER FILING DATE: 1999-10-21
; EARLIER FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 517
; TYPE: PRT
; ORGANISM: Human
US-09-422-869-6

Query Match 74.4%; Score 2715; DB 5; Length 517;
Best Local Similarity 76.8%; Pred. No. 2e-264;
Matches 516; Conservative 1; Mismatches 0; Indels 155; Gaps 1;

```

QY 1 MRAGGATPARELFRDAAPPAADSSLCFCDLSTPLAQFREDITWRRPOICATPRLFPDDP 60
Db 1 MRAGGATPARELFRDAAPPAADSSLCFCDLSTPLAQFREDITWRRPOICATPRLFPDDP 60
QY 61 REGOVKOGLLGDCWFLCACAALOKSRHLLDOVIPPGOPSWADQYRGSGFTCRIMQFGRWV 120
Db 61 REGOVKOGLLGDCWFLCACAALOKSRHLLDOVIPPGOPSWADQYRGSGFTCRIMQFGRWV 120
QY 121 EVTTDDRLPCLAGRLCFSCRCQREDFVFWLPLEKRYAARHSGSYEHLMAGVADALVDLTGG 180
Db 121 EVTTDDRLPCLAGRLCFSCRCQREDFVFWLPLEKRYAARHSGSYEHLMAGVADALVDLTGG 180
QY 121 EVTTDDRLPCLAGRLCFSCRCQREDFVFWLPLEKRYAARHSGSYEHLMAGVADALVDLTGG 180
Db 121 EVTTDDRLPCLAGRLCFSCRCQREDFVFWLPLEKRYAARHSGSYEHLMAGVADALVDLTGG 180
QY 181 LAERWNLKGVAGSGGQODRPGRMWHRTRCROLLHLKDQCLISCCVLSFRAGARELGEFHA 240
Db 181 LAERWNLKGVAGSGGQODRPGRMWHRTRCROLLHLKDQCLISCCVLSFRAGARELGEFHA 240
QY 241 IVSDRLRELQAGOCILLRIQNPWGRRCWGLMRREGEGMSQYDAVASSELLSQLQEGE 300
Db 241 IVSDRLRELQAGOCILLRIQNPWGRRCWGLMRREGEGMSQYDAVASSELLSQLQEGE 300
QY 301 FWVEEERFLREFDELFTGYPTVTEAGHLOSILYTERLCHTRALPGAWKGGAGGCRNNG 360
Db 301 FWVEEERFLREFDELFTGYPTVTEAGHLOSILYTERLCHTRALPGAWKGGAGGCRNNG 360
QY 361 FPSNPKFWLVSSEPEYIYAVLQSRSLHAADWAGARALVGDSTMSPASISGKHQAV 420
Db 361 FPSNPKFWLVSSEPEYIYAVLQSRSLHAADWAGARALVGDSTMSPASISGKHQAV 420
QY 421 GLHLMKVEKRRVNLPRVLSMPVAGTACHAYDREVMHRCESLSPGYIYAVSTELKDAPEG 480
Db 421 GLHLMKVEKRRVNLPRVLSMPVAGTACHAYDREVMHRCESLSPGYIYAVSTELKDAPEG 480
QY 481 FLIRVSTGRVSLALRAVAKNTTGAALPAGENGTVOLGSRWVGQTAGSRNFASTPT 540
Db 481 FLIRVSTGRVSLALRAVAKNTTGAALPAGENGTVOLGSRWVGQTAGSRNFASTPT 540
QY 541 NPCEPSPVPEGPGRVCRTILHQCRCPSDTEHPHIGFHIQVPEGGRSODAPRLLOEPL 600
Db 541 NPCEPSPVPEGPGRVCRTILHQCRCPSDTEHPHIGFHIQVPEGGRSODAPRLLOEPL 600
QY 601 LSCVPRVAQEVSRCLLPAGTYKYVPSTYLPDTEGAFYTIATRIIDRPSIHQEMIGOP 660
Db 601 LSCVPRVAQEVSRCLLPAGTYKYVPSTYLPDTEGAFYTIATRIIDRPSIHQEMIGOP 660
QY 661 LOEVSVMAVWKT 672
Db 661 LOEVSVMAVWKT 672
QY 506 LOEVSVMAVWKT 517
Db 506 LOEVSVMAVWKT 517

```

```

; LENGTH: 544
; TYPE: PRT
; ORGANISM: Human
US-09-422-869-4

Query Match
Best Local Similarity 74.0%; Score 2701.5; DB 4; Length 544;
Matches 498; Conservative 5; Mismatches 6; Indels 13; Gaps 1;

QY 1 MRAGGATPARELFRDAAPPAADSSLCFCDLSTPLAQFREDITWRRPOICATPRLFPDDP 60
Db 1 MRAGGATPARELFRDAAPPAADSSLCFCDLSTPLAQFREDITWRRPOICATPRLFPDDP 60
QY 61 REGOVKOGLLGDCWFLCACAALOKSRHLLDOVIPPGOPSWADQYRGSGFTCRIMQFGRWV 120
Db 61 REGOVKOGLLGDCWFLCACAALOKSRHLLDOVIPPGOPSWADQYRGSGFTCRIMQFGRWV 120
QY 121 EVTTDDRLPCLAGRLCFSCRCQREDFVFWLPLEKRYAARHSGSYEHLMAGVADALVDLTGG 180
Db 121 EVTTDDRLPCLAGRLCFSCRCQREDFVFWLPLEKRYAARHSGSYEHLMAGVADALVDLTGG 180
QY 181 LAERWNLKGVAGSGGQODRPGRMWHRTRCROLLHLKDQCLISCCVLSFRAGARELGEFHA 240
Db 181 LAERWNLKGVAGSGGQODRPGRMWHRTRCROLLHLKDQCLISCCVLSFRAGARELGEFHA 240
QY 241 IVSDRLRELQAGOCILLRIQNPWGRRCWGLMRREGEGMSQYDAVASSELLSQLQEGE 300
Db 241 IVSDRLRELQAGOCILLRIQNPWGRRCWGLMRREGEGMSQYDAVASSELLSQLQEGE 300
QY 301 FWVEEERFLREFDELFTGYPTVTEAGHLOSILYTERLCHTRALPGAWKGGAGGCRNNG 360
Db 301 FWVEEERFLREFDELFTGYPTVTEAGHLOSILYTERLCHTRALPGAWKGGAGGCRNNG 360
QY 361 FPSNPKFWLVSSEPEYIYAVLQSRSLHAADWAGARALVGDSTMSPASISGKHQAV 420
Db 361 FPSNPKFWLVSSEPEYIYAVLQSRSLHAADWAGARALVGDSTMSPASISGKHQAV 420
QY 421 GLHLMKVEKRRVNLPRVLSMPVAGTACHAYDREVMHRCESLSPGYIYAVSTELKDAPEG 480
Db 421 GLHLMKVEKRRVNLPRVLSMPVAGTACHAYDREVMHRCESLSPGYIYAVSTELKDAPEG 480
QY 481 FLIRVSTGRVSLALRAVAKNTTGAALPAGENGTVOLGSRWVGQTAGSRNFASTPT 540
Db 481 FLIRVSTGRVSLALRAVAKNTTGAALPAGENGTVOLGSRWVGQTAGSRNFASTPT 540

```

```

RESULT 5
US-09-422-869-8
; Sequence 8, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: ODA, NAOHISA
; APPLICANT: COX, NANCY J.
; APPLICANT: SREENAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: HANIS, KENICHI
; APPLICANT: BEIL, GRAEME I.
; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
; FILE REFERENCE: ARCD:307
; CURRENT FILING DATE: 1999-10-21
; EARLIER APPLICATION NUMBER: 60/134,175
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 8
; LENGTH: 513
; TYPE: PRT
; ORGANISM: Human

```

US-09-422-869-8

Query Match 74.0%; Score 2699; DB 4; Length 513;
Best Local Similarity 98.2%; Pred. No. 8.9e-263;
Matches 496; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 1 MRAGGATPARELFDAAFPAAADSSLCFCDLSTPLAQFREDITWRRPOEICATPRLFPDDP 60
DB 1 MRAGGATPARELFDAAFPAAADSSLCFCDLSTPLAQFREDITWRRPOEICATPRLFPDDP 60

QY 61 REGOVKGLGDCWFLCACAALQKSRHLQDVIPPGQPSWADOEYRGSTCRINQFGRW 120
DB 61 REGOVKGLGDCWFLCACAALQKSRHLQDVIPPGQPSWADOEYRGSTCRINQFGRW 120

QY 121 EVTTDDRLPCLAGRLCFRCQREDVFWLPLEKVKYAKVHGSYEHLMWAGOVADALVDLTGG 180
DB 121 EVTTDDRLPCLAGRLCFRCQREDVFWLPLEKVKYAKVHGSYEHLMWAGOVADALVDLTGG 180

QY 181 LAERNLKGAVSGGQODRPGRWEHRTCRQLHLKDOCLISCCVLSPRAGARELGEHAF 240
DB 181 LAERNLKGAVSGGQODRPGRWEHRTCRQLHLKDOCLISCCVLSPRAGARELGEHAF 240

QY 241 IVSDRLQOGAGOCILLRIONPWRCWQGLWREGGEGSWQDAVAASELLSLOQEGE 300
DB 241 IVSDRLQOGAGOCILLRIONPWRCWQGLWREGGEGSWQDAVAASELLSLOQEGE 300

QY 301 FWEEEEFLREFDELTVGYPVTEAGHLQSLYTERLLCHTRALPGAWKQSGAGCRNNSG 360
DB 301 FWEEEEFLREFDELTVGYPVTEAGHLQSLYTERLLCHTRALPGAWKQSGAGCRNNSG 360

QY 361 FPSNPKFWLRVSEPVYIAVLRSLHAADWAGARALVGDSTHSWSPASIPCKHYQAV 420
DB 361 FPSNPKFWLRVSEPVYIAVLRSLHAADWAGARALVGDSTHSWSPASIPCKHYQAV 420

QY 421 GLHLWKVKRNVLPVLSNPPVAGTACHAYDREHVLRCLESPGYLLAVPSTFLKDPAGE 480
DB 421 GLHLWKVKRNVLPVLSNPPVAGTACHAYDREHVLRCLESPGYLLAVPSTFLKDPAGE 480

QY 481 FLLRVSTGRVSLSAIRAVAKNTTP 505
DB 481 FLLRVSTGRVSLRSQVEGARTHP 505

RESULT 6

US-09-422-869-10
; Sequence 10, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: ODA, NAOHISA
; APPLICANT: COX, NANCY J.
; APPLICANT: SREENAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: OTANI, KENICHI
; APPLICANT: HANIS, CRAIG L.
; APPLICANT: BELL, GRAEME I.
; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
; FILE REFERENCE: ARCD:307
; CURRENT APPLICATION NUMBER: US/09/422,869
; CURRENT FILING DATE: 1999-10-21
; EARLIER APPLICATION NUMBER: 60/134,175
; EARLIER FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 444
; TYPE: PRT
; ORGANISM: Human
US-09-422-869-10

Query Match 64.0%; Score 2336; DB 4; Length 444;
Best Local Similarity 99.8%; Pred. No. 2.3e-226;
Matches 425; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRAGGATPARELFDAAFPAAADSSLCFCDLSTPLAQFREDITWRRPOEICATPRLFPDDP 60
DB 1 MRAGGATPARELFDAAFPAAADSSLCFCDLSTPLAQFREDITWRRPOEICATPRLFPDDP 60

QY 61 REGOVKGLGDCWFLCACAALQKSRHLQDVIPPGQPSWADOEYRGSTCRINQFGRW 120
DB 61 REGOVKGLGDCWFLCACAALQKSRHLQDVIPPGQPSWADOEYRGSTCRINQFGRW 120

QY 121 EVTTDDRLPCLAGRLCFRCQREDVFWLPLEKVKYAKVHGSYEHLMWAGOVADALVDLTGG 180
DB 121 EVTTDDRLPCLAGRLCFRCQREDVFWLPLEKVKYAKVHGSYEHLMWAGOVADALVDLTGG 180

QY 181 LAERNLKGAVSGGQODRPGRWEHRTCRQLHLKDOCLISCCVLSPRAGARELGEHAF 240
DB 181 LAERNLKGAVSGGQODRPGRWEHRTCRQLHLKDOCLISCCVLSPRAGARELGEHAF 240

QY 241 IVSDRLQOGAGOCILLRIONPWRCWQGLWREGGEGSWQDAVAASELLSLOQEGE 300
DB 241 IVSDRLQOGAGOCILLRIONPWRCWQGLWREGGEGSWQDAVAASELLSLOQEGE 300

QY 301 FWEEEEFLREFDELTVGYPVTEAGHLQSLYTERLLCHTRALPGAWKQSGAGCRNNSG 360
DB 301 FWEEEEFLREFDELTVGYPVTEAGHLQSLYTERLLCHTRALPGAWKQSGAGCRNNSG 360

QY 361 FPSNPKFWLRVSEPVYIAVLRSLHAADWAGARALVGDSTHSWSPASIPCKHYQAV 420
DB 361 FPSNPKFWLRVSEPVYIAVLRSLHAADWAGARALVGDSTHSWSPASIPCKHYQAV 420

QY 421 GLHLWK 426
DB 421 GLHLWK 426

RESULT 7

US-09-422-869-12
; Sequence 12, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: ODA, NAOHISA
; APPLICANT: COX, NANCY J.
; APPLICANT: SREENAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: OTANI, KENICHI
; APPLICANT: HANIS, CRAIG L.
; APPLICANT: BELL, GRAEME I.
; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
; FILE REFERENCE: ARCD:307
; CURRENT APPLICATION NUMBER: US/09/422,869
; CURRENT FILING DATE: 1999-10-21
; EARLIER APPLICATION NUMBER: 60/134,175
; EARLIER FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 274
; TYPE: PRT
; ORGANISM: Human
US-09-422-869-12

Query Match 23.5%; Score 858.5; DB 4; Length 274;
Best Local Similarity 81.7%; Pred. No. 4.6e-78;
Matches 165; Conservative 1; Mismatches 13; Indels 23; Gaps 3;

QY 1 MRAGGATPARELFDAAFPAAADSSLCFCDLSTPLAQFREDITWRRPOEICATPRLFPDDP 60
DB 1 MRAGGATPARELFDAAFPAAADSSLCFCDLSTPLAQFREDITWRRPOEICATPRLFPDDP 60

QY 61 RESOVKGLIGDCWFLCAGALOKSRHLLDOYIPPCGSPWADQYRGSTCRIMQGRMV 120
 DB 61 RESOVKGLIGDCWFLCAGALOKSRHLLDOYIPPCGSPWADQYRGSTCRIMQGRMV 120
 QY 121 EYTTDDRLPCLAGRLCFSCQREDVFWPLLEKVVAKVHVSYEHLMAGVADALVDLTG 180
 DB 121 EYTTDDRLPCLAGRLCFSCQREDVFWPLLEKVVAKVHVSYEHLMAGVADALVDLTG 180
 QY 181 LAERNMLKAVAGSGGQDDRRGR 202
 DB 168 -----GCPGGRDRRGR 179

RESULT 8

US-08-835-099A-1
 ; Sequence 1, Application US/08835099A
 ; Patent No. 5874277
 ; GENERAL INFORMATION:
 ; APPLICANT: SHINTANI, Yasushi
 ; APPLICANT: NISHI, Kazuhiro
 ; APPLICANT: KAWAMOTO, Tomohiro
 ; TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION
 ; TITLE OF INVENTION: AND USE
 ; NUMBER OF SEQUENCES: 18
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
 ; STREET: 130 Water Street
 ; CITY: Boston
 ; STATE: MA
 ; COUNTRY: USA
 ; ZIP: 02109
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FASTSEQ for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/835, 099A
 ; FILING DATE: 04-APR-1997
 ; CLASSIFICATION: 514
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 083649/1996
 ; FILING DATE: 05-APR-1996
 ; APPLICATION NUMBER: 97105508.2
 ; FILING DATE: 03-APR-1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Resnick, David S
 ; REGISTRATION NUMBER: 34,235
 ; REFERENCE/DOCKET NUMBER: 47342
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 617-523-3400
 ; TELEFAX: 617-523-6440
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 703 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-835-099A-1

Query Match 17.4%; Score 635; DB 2; Length 703;
 Best local Similarity 32.3%; Pred. No. 6, 5e-55;

Matches 164; Conservative 85; Mismatches 179; Indels 80; Gaps 17;
 QY 13 LFRDAFPADSSLC-FCDLSTPLAQFREDITWRROECATPRLFPDDPREGVKQGLG 71
 DB 45 LFRDPEPPACPSALGKYDGLG-PGSPOTQGIIMKRPTELCSPGYIGAGARTIDICQGLG 103
 QY 72 DCNFIACALOKSRHLLDOYIPPCGSPWADQYRGSTCRIMQGRMV EYTTDDRLPCL 131

DB 104 DCMILAAATSLINELLILRYVRPDQ--FDENYAGITHFQWQYGEVVEVYIDRLPTK 161
 QY 132 AGRCLFSCQREDVFWPLLEKVVAKVHVSYEHLMAGVADALVDLTGGLAERNMLKQVA 191
 DB 162 NGQLFLHSBQENEFMSALLEKAYKINGCYEALAGSTVGEFEDFTGTSISFDLK--- 218
 QY 192 GSGGQDDRRGRHRTCCQLHLKQDC---LISCCV-LSPRAGA-----RELGEFHATV 242
 DB 219 -----KPRANLYOITIKAL-----CAGSLGCSIDVSAAEATISOKLYSHAYSV 266
 QY 243 SDLEIQAQAGCILLIRIONPMGRQWGLMREGSGMSOVDAVAVSELSLOLQEGEFW 302
 DB 267 TGVEEVNQ-GHEPLILRLNPMGVEVMSGAMSDAPFMNHIDRRKEELKQVDEGEFW 325
 QY 303 VEEEFLEPEDELVGYVTENAGHLQSLYTERLLCHTRALPGAWYKQSGAGCRNNSGFP 362
 DB 326 MSLSDVFRQFRLLEICNLSPDSLSEEVHKNMVLFN---GHWTRGSTAGGCON---YP 378
 QY 363 ----SNPKFWLRVSEPS-----YIYAVLQSRILHADMAGRALYDGSHTS 406
 DB 379 ATYTNPQFKIRLDEVDDEESIGECCTVLLGIMQNR-----RMKRRRIGQGMIS 430
 QY 407 WSPASIPGKHVQAVGLHLMKYEKR-----RVNLP--VLSMPVAGTACHAYDREYHLRC 459
 DB 431 -----IGYAVYQPKLESHTDHLGRDFLAFQPSARTSTYVNLKEVSGRA 477
 QY 460 ELSPGYIYAVSTFLKADPGEFLLRVPS 487
 DB 478 RLPPGEYIYVSPSTPEPDEGEFLRVPS 505

RESULT 9

US-09-157-349-1
 ; Sequence 1, Application US/09157349
 ; Patent No. 6068990
 ; GENERAL INFORMATION:
 ; APPLICANT: SHINTANI, Yasushi
 ; APPLICANT: NISHI, Kazuhiro
 ; APPLICANT: KAWAMOTO, Tomohiro
 ; TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION
 ; TITLE OF INVENTION: AND USE
 ; NUMBER OF SEQUENCES: 18
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
 ; STREET: 130 Water Street
 ; CITY: Boston
 ; STATE: MA
 ; COUNTRY: USA
 ; ZIP: 02109
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FASTSEQ for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/157, 349
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/835, 099
 ; FILING DATE:
 ; APPLICATION NUMBER: 97105508.2
 ; FILING DATE: 03-APR-1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Resnick, David S
 ; REGISTRATION NUMBER: 34,235
 ; REFERENCE/DOCKET NUMBER: 47342
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 617-523-3400
 ; TELEFAX: 617-523-6440
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 703 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-157-349-1

Query Match 17.4%; Score 635; DB 3; Length 703;
Best Local Similarity 32.3%; Pred. No. 6.5e-55;
Matches 164; Conservative 85; Mismatches 179; Indels 80; Gaps 17;

QY 13 LFRDAAPPAADSS-L-FCDLSTPLAQFREDITWRRPOEICATPRLFPDDPREQVKGGLG 71
DB 45 LFKDFEPFACPSALGYKDLG-PCSPOTOGIWKRTPELCPSPQFIVGGATRDICQGG 103
QY 72 DCWFLCACAALOKSHLLDQVIPPQPSWADQYEGSFTCRWQGRWVEVTTDRRLCL 131
DB 104 DCWLLAAIASLNLNEELLYRVPRDQ--FOENYAGIFHFQWQYGEWVEVVDRLPTK 161
QY 132 AGRLCFSQCRQEDVFWLPLEKVKYAKVHGSYHLMWAGQVADALVDLTGGLAERWNLKGYA 191
DB 162 NGOLLFLHSEQGNFWSALLEKAYAKNGCYEALAGGSTVEGFEDTGGISEFYDLK--- 218
QY 192 GSGGQDPRGRWEHTRCQLLHLKQDQ---LISCCV-LSPRAGA-----RELGEFHAFIV 242
DB 219 -----KPPANLYQIIRKAL-----CAGSLGCSIDVSSAAEAIAITSQKLKSHAYSV 266
QY 243 SDRLEQOAGQCILLRIQNPWGRRCWQGLWREGGEGSQVDAVASSELLSQLOEGEFW 302
DB 267 TGVVEVNFQ-GHPEKLIIRLNPWGEVWSGAWSDDAPEWNIIDPRRKEELDKKVEDGEFW 325
QY 303 VEEEFLEFDELTVGYPVTEAGHLQSLYTERLLCHTRALPCAWKVGOSAGCGRNNSGFP 362
DB 326 MSLSDFVRQFSRLICNLSPDLSSEEVHKWNLVFN---GHWTRGSTAGCQCN---YP 378
QY 363 -----SNPKFWLRVSEPE-----VYIAVLRSLRHAADWAGRALVGDSTHS 406
DB 379 ATYTNPFQKIRLDEVEDQESIGEPCTVLLGLMOKNR-----RWRKRIGQGML 430
QY 407 WSPASIPKHYQAVGLHLWKVEKR-----RVNLPR--VLSMPPVAGTACHAYDREVHLRC 459
DB 431 -----IGYAVYQVPKELESHTDAHLGRDFFLAYQPSARTSTYVNLREVSGRA 477
QY 460 ELSPGYLVAVPSTFLKDPAGEFLLRVFS 487
DB 478 RLPPGEYLVPSTFEPFKDGEFCFLRVFS 505

RESULT 10
US-08-835-099A-2
Sequence 2, Application US/08835099A
Patent No. 5874277
GENERAL INFORMATION:
APPLICANT: SHINTANI, Yasushi
APPLICANT: NISHI, Kazuori
APPLICANT: KAWAMOTO, Tomohiro
TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION
TITLE OF INVENTION: AND USE
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/835,099A
FILING DATE: 04-APR-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 083649/1996
FILING DATE: 05-APR-1996
APPLICATION NUMBER: 97105508.2
FILING DATE: 03-APR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Resnick, David S
REGISTRATION NUMBER: 34,235
REFERENCE/DOCKET NUMBER: 47342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 712 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-835-099A-2

Query Match 17.4%; Score 635; DB 2; Length 712;
Best Local Similarity 32.3%; Pred. No. 6.6e-55;
Matches 164; Conservative 85; Mismatches 179; Indels 80; Gaps 17;

QY 13 LFRDAAPPAADSS-L-FCDLSTPLAQFREDITWRRPOEICATPRLFPDDPREQVKGGLG 71
DB 54 LFKDFEPFACPSALGYKDLG-PCSPOTOGIWKRTPELCPSPQFIVGGATRDICQGG 112
QY 72 DCWFLCACAALOKSHLLDQVIPPQPSWADQYEGSFTCRWQGRWVEVTTDRRLCL 131
DB 113 DCWLLAAIASLNLNEELLYRVPRDQ--FOENYAGIFHFQWQYGEWVEVVDRLPTK 170
QY 132 AGRLCFSQCRQEDVFWLPLEKVKYAKVHGSYHLMWAGQVADALVDLTGGLAERWNLKGYA 191
DB 171 NGOLLFLHSEQGNFWSALLEKAYAKNGCYEALAGGSTVEGFEDTGGISEFYDLK--- 227
QY 192 GSGGQDPRGRWEHTRCQLLHLKQDQ---LISCCV-LSPRAGA-----RELGEFHAFIV 242
DB 228 -----KPPANLYQIIRKAL-----CAGSLGCSIDVSSAAEAIAITSQKLKSHAYSV 275
QY 243 SDRLEQOAGQCILLRIQNPWGRRCWQGLWREGGEGSQVDAVASSELLSQLOEGEFW 302
DB 276 TGVVEVNFQ-GHPEKLIIRLNPWGEVWSGAWSDDAPEWNIIDPRRKEELDKKVEDGEFW 334
QY 303 VEEEFLEFDELTVGYPVTEAGHLQSLYTERLLCHTRALPCAWKVGOSAGCGRNNSGFP 362
DB 335 MSLSDFVRQFSRLICNLSPDLSSEEVHKWNLVFN---GHWTRGSTAGCQCN---YP 387
QY 363 -----SNPKFWLRVSEPE-----VYIAVLRSLRHAADWAGRALVGDSTHS 406
DB 388 ATYTNPFQKIRLDEVEDQESIGEPCTVLLGLMOKNR-----RWRKRIGQGML 439
QY 407 WSPASIPKHYQAVGLHLWKVEKR-----RVNLPR--VLSMPPVAGTACHAYDREVHLRC 459
DB 440 -----IGYAVYQVPKELESHTDAHLGRDFFLAYQPSARTSTYVNLREVSGRA 486
QY 460 ELSPGYLVAVPSTFLKDPAGEFLLRVFS 487
DB 487 RLPPGEYLVPSTFEPFKDGEFCFLRVFS 514

RESULT 11
US-09-157-349-2
Sequence 2, Application US/09157349
Patent No. 6068990
GENERAL INFORMATION:

Fri Sep 20 10:47:44 2002

us-09-768-877-2.ra1

Page 7

APPLICANT: SHINTANI, Yasushi
APPLICANT: NISHI, Kazuori
APPLICANT: KAWAMOTO, Tomohiro
TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION
TITLE OF INVENTION: AND USE
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESS: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/157,349
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/835,099
FILING DATE:
APPLICATION NUMBER: 97105508.2
FILING DATE: 03-APR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Resnick, David S
REGISTRATION NUMBER: 34,235
REFERENCE/DOCKET NUMBER: 47342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 712 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-157-349-2

Query Match 17.4% Score 635; DB 3; Length 712;
Best Local Similarity 32.3%; Pred. No. 6,6e-55;
Matches 164; Conservative 85; Mismatches 179; Indels 80; Gaps 17;
QY 13 LFRDAFPAADSL-FCDLSTPLAOFREDITWRPQICATPLFPDDPREGQKGLG 71
DB 54 LKDEFPACPSALGKDLG-PGSPOTGCIIMKRPTELCPFOIYVGATRDICGGG 112
QY 72 DCFWLCACAAALOKSRHLDOVTPPQPSWADERYGSCFCRIWQFGWVEVTTDRLPCL 131
DB 113 DCFWLCACAAALOKSRHLDOVTPPQPSWADERYGSCFCRIWQFGWVEVTTDRLPCL 170
QY 132 AGRLECSRCORREDVFWPLPLEKVAKVHSGSYELHNAQYADLVDTGLAERMLKGYA 191
DB 171 NQDLFLHSEQNEFWMSALLEKAVAKLNGCYALAGSGVSEGFEDFTGGSISEFYDLK 227
QY 192 GSGGQODRPGRWEHRTCRQLHLKQDC--LISCCV-LSPRAGA-----RELGEFPAFV 242
DB 228 -----KPRANLYOIIRKAL-----CAGSLGCSIDVSSAAEAELITISQKLVKSHAVS 275
QY 243 SDRELQAGACICILLRIQNPWGRRCWGLMREGGEGNSQYDAVAASSELLSQLEGEFWE 302
DB 276 TGVEEVNPO-GHPEKILIRLNPWGEVEMSGAUSDAPENHIDPRKEELDKKVDGEFWE 334
QY 303 VEEBEFLREFDELIVGYPTENGHLSLYTERLCHTRALPGAWYKGSAGCCRNNSGFP 362
DB 335 MSLSDVRFKSLTETCNLSPSLSSSEVHNKMLVLEF--GHWTGSGTAGCCQ--YF 387

QY 363 -----SNPKFLRVSESE-----VYIAVLORSRLHAADWAGARALVDSHTS 406
DB 368 ATWTNPQFKIRLDEVEDDEESIGECCTVLGLMOKNR-----RRKRIGQCMLS 439
QY 407 MSPASIPGKHVQAVGLHMLVEKR-----RVNLP--VLSMPYAGTACIAYREYHLC 459
DB 440 -----IGAVYVPEKLESHDHAHGRDFLAVOPSARTSYVNLREVSRA 486
QY 460 ELSPGYLAVPSTFLKADAGEFLRVFS 487
DB 487 RLPGGYLVVSTPEPKEDEFLRVFS 514
RESULT 12
US-09-422-869-22
Sequence 22, Application US/09422869
Patent No. 6235481
GENERAL INFORMATION:
APPLICANT: POLONSKY, KENNETH S.
APPLICANT: HORIKAWA, YUKIO
APPLICANT: ODA, NAOHISA
APPLICANT: COX, NANCY J.
APPLICANT: SEENAN, SEAMUS
APPLICANT: ZHOU, YUN-PING
APPLICANT: OTANI, KENICHI
APPLICANT: HANIS, CRAIG L.
APPLICANT: BELL, GRAEME I.
TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
FILE REFERENCE: ARCD:307
CURRENT APPLICATION NUMBER: US/09/422,869
CURRENT FILING DATE: 1999-10-21
EARLIER APPLICATION NUMBER: 60/134,175
EARLIER FILING DATE: 1999-05-13
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 22
LENGTH: 714
TYPE: prp
ORGANISM: Human
US-09-422-869-22

Query Match 16.9% Score 618.5; DB 4; Length 714;
Best Local Similarity 32.3%; Pred. No. 3e-53;
Matches 166; Conservative 73; Mismatches 198; Indels 77; Gaps 17;

QY 13 LFRDAFPAADSL-FCDLSTPLAOFREDITWRPQICATPLFPDDPREGQKGLG 71
DB 55 LFRDAFPAADSL-FCDLSTPLAOFREDITWRPQICATPLFPDDPREGQKGLG 113
QY 72 DCFWLCACAAALOKSRHLDOVTPPQPSWADERYGSCFCRIWQFGWVEVTTDRLPCL 131
DB 114 DCFWLCACAAALOKSRHLDOVTPPQPSWADERYGSCFCRIWQFGWVEVTTDRLPCL 171
QY 132 AGRLECSRCORREDVFWPLPLEKVAKVHSGSYELHNAQYADLVDTGLAERMLKGYA 191
DB 172 DCFWLCACAAALOKSRHLDOVTPPQPSWADERYGSCFCRIWQFGWVEVTTDRLPCL 231
QY 192 GSGGQODRPGRWEHRTCRQLHLKQDC--LISCCV-LSPRAGA-----RELGEFPAFV 245
DB 232 SDLYO-----LTKALERSLGCSDISSVYDMEALITFKLVKHAASVTGA 279
QY 246 RELQAGACICILLRIQNPWGRRCWGLMREGGEGNSQYDAVAASSELLSQLEGEFWE 305
DB 280 KQV-NYRGQVSLIRBNPWGEVEMSGAUSDAPENHIDPRKEELDKKVDGEFWE 338
QY 306 EEFLEFDELIVGYPTENGHLSLYTERLCHTRALPGAWYKGSAGCCRNNSGFP 363
DB 339 RQFMEFDELIVGYPTENGHLSLYTERLCHTRALPGAWYKGSAGCCRNNSGFP 391
QY 364 -----NPKFVLVSESE-----VYIAVLORSRLHAADWAGARALVDSHTS 409
DB 392 WYNPQFKIRLDEVEDDEESIGECCTVLGLMOKNR-----RRKRIGQCMLS 434

```

Qy 410 ASTPGKHQYQAVGLHLWKVKERRNNLPV -LSMPVPVAGTACHAYD-----REVHLRCELS 462
      : : : : :
Db 435 -----GRDMETIGFAVTEYVELVQGPVAVHLKRDFFLANASRSEOFINLREVSTFRFLP 490
      : : : : :
Qy 463 PGYLAVPSTFLKDDAPGCEFLRVF-----STGRVSL 493
      : : : : :
Db 491 PGYIVVVPSTFEPNKGEGDFVLRFFSEKKSAGTVEL 524
      : : : : :

```

```

RECORD 13
US-09-422-869-24
; Sequence 24, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: ODA, NAOHISA
; APPLICANT: COX, NANCY J.
; APPLICANT: GREENAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: OTANI, KENICHI
; APPLICANT: HANIS, CRAIG L.
; APPLICANT: BELL, GRAEME I.
; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
; FILE REFERENCE: ARCD:307
; CURRENT APPLICATION NUMBER: US/09/422,869
; CURRENT FILING DATE: 1999-10-21
; EARLIER APPLICATION NUMBER: 60/134,175
; EARLIER FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO. 24
; LENGTH: 821
; TYPE: PRT
; ORGANISM: Human
US-09-422-869-24

```

```

Query Match      16.9%; Score 618; DB 4; Length 821;
Best Local Similarity 31.4%; Pred. No. 4.3e-53;
Matches 178; Conservative 79; Mismatches 194; Indels 116; Gaps 19;

13 LFDDAAFPADSSLFCDLSPLAQFREDIWRBPQETACPRLPDPDPRGCVKQGILGD 72
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
14 LYVDPEPPDETSLFSQKEPI-QF-----VWKRPPEICEPRFTIDGANTDTICOGELGD 128
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
73 CWFCLCAALOKSRHLLDVIPPGQPSWADOEYRGSTCRIFOWGRVVEVTTDRLEPCLA 132
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
129 CWFLLAATLTLNOHLFLRVPHPDQS-FIENYAGIFHFQFWRYGEWVDVVDDCLPTYN 186
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
133 GRLCFSRCOREDVFVPLEKKYAKVHGVSYEHLWAGOVADALVDLTGGLA----- 182
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
187 NQLVFTKSNHRNFEWSALLEKAYALKHGSYEAALKGGNTTEAMEDFTGVGAFFEIFRDAPS 246
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
183 -----ERNWLKGVA-----GSGGQQDRGRWEHRTCROLLHKQCLISCC 223
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
247 DMVKINKKAIEGSLMGCSIDDGTNYTGYSPSGLANGELIARVRNN-----DNSLLQDS 302
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
224 VLSFRAG-----ARELGEFHAFTVSDRLERLAQQAQCILLIRIQNPWG 266
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
303 DLDPRGSDERPRTITIPVOYETRMACGLVRGHAYSVTGLDEVPEK-GEKVLYRLRNPMWG 361
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
267 RRCWMOGLRWEGGEGNSQVDAAVASELLSQL-QGEFVFWEEEEEFLREFDELTUVGPVTFEAG 325
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
362 QVENNGSWSDRWKDMSFDVKDKEXARLOHQVTEDEGFWMYSYEDFIYHFTKLEICNLTA-DA- 420
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
326 HLOSLXTERLLCHLTRAL-PCGAWKYGOSAGGCCRRNSGFP-----SNPKFMRVSE-----p 374
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
421 ----LQSDKLOTWTVSVNEGRVWVGCSAGGCNRN---FPDTFTWNQPTRLKLEDDDDDDDD 473
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
375 SEV-----YIAVLQRSLHAADWAGRALVCGDSHTSWSPASIPCKHYQAVGLHLHWKY---- 427
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db      474  SEVICSEFLVALMQRNR-----RKDRKL-----GASLFTIGFAIYEVPKR 512

Qy      428  ---EKRRVNLPRVLSPMPPVAGTACHAYDREHLCRLCELSPGYYLAVPSTFELKDAPGEFLLR 484
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      513  MHGNKQHLOKDFEFLYNASKARSKTYINMRREVYSQRFLPPSEVVIVPSTYEPHQBGEFILR 572
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy      485  VFSTR-----VSLSAIRAVAKNTT 504
           ||| | : : | | | |
Db      573  VFSEKRNLSEVENTISVDPRVKKKKT 599
           : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
US-09-422-869-27
; Sequence 27, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: ODA, NAOHISA
; APPLICANT: COX, NANCY J.
; APPLICANT: SREENAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: OTANI, KENICHI
; APPLICANT: HANIS, CRAIG L.
; APPLICANT: BELL, GRAEME I.
; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
; FILE REFERENCE: ARCD:307
; CURRENT APPLICATION NUMBER: US/09/422,869
; CURRENT FILING DATE: 1999-10-21
; EARLIER APPLICATION NUMBER: 60/134,175
; EARLIER FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 27
; LENGTH: 703
; TYPE: PRT
; ORGANISM: RAT
; US-09-422-869-27

```

```

Query Match      16.88; Score 612; DB 4; Length 703;
Best Local Similarity 31.2%; Pred. No. 1.3e-52;
Matches 155; Conservative 87; Mismatches 197; Indels 58; Gaps 14;

QY      13  LFRDAAPPAADSSLL-FCDLSTPLAOFREDITWRBPQEQICATPLRFPDPREGQVKQGLG 71
      Db      111 111 111 : 111 : 111 : 111 111 111 : 111 111 : 111 111
QY      72  DCWFPLCAALQKSRHLLDQVIPPQGPQSWADQYVRSFTCRIMQFGRWVEYTTDRRLPCL 131
      Db      111 111 111 : 111 : 111 : 111 111 111 : 111 111 : 111 111
QY      104 DCWLLAAIASLTLNEKLLRYRLPRDQS--FKQDYAGIFHFQFQWGYGEVWVWIDRLPTK 161
      Db      111 111 111 : 111 : 111 : 111 111 111 : 111 111 : 111 111
QY      132 AGRLCFSRCOREDVEWLPLEKLYYAKVGHVSVEHLWAGQADALVDLNGGLAERNLKGVA 191
      Db      111 111 111 : 111 111 111 111 111 111 : 111 111 : 111 111
QY      162 NGQLLFHSHSEGFNEFWSALKEKAYKALGSTTEALVGSTIEGFEDFTGGSEFYDLK--- 218
      Db      111 111 111 : 111 111 111 111 111 111 : 111 111 : 111 111
QY      192 GSGGQDPRGRWEHRTCRQLLHLKQDLISCCV--LSPRAGA-----RELGEHAFIVSDL 245
      Db      111 111 111 : 111 : 111 : 111 111 111 : 111 111 : 111 111
QY      219 -----KPPENLYYIIQKA--LRKGSLLGCSIDVSTAAEAETTRQKLVKGHAYSVTG 269
      Db      111 111 111 : 111 : 111 : 111 111 111 : 111 111 : 111 111
QY      246 RELQGAQCILLIRIONPGRRCRWCGLWREGGEGNSQVDAAVASELLSLOQEGEFWEE 305
      Db      111 111 111 : 111 : 111 : 111 111 111 : 111 111 : 111 111
QY      270 EEFNFH--GRPEKILRLRNPWGEVWSGAWSDNAPEWYIDPRKKEELDCKAEDGEFWMSE 328
      Db      111 111 111 : 111 : 111 : 111 111 111 : 111 111 : 111 111
QY      306 EEFLEFDELTVGVPVTEAGHLQSLYTERLLICHTRALPAGWAKGASQAGCSNNNS--FPSN 364
      Db      111 111 111 : 111 : 111 : 111 111 111 : 111 111 : 111 111
QY      329 SDFLKQYSREICNLSPDLSLSEEHKWNVLFN-----GRWTRGSTAGGCLNTPGTTWTN 384
      Db      111 111 111 : 111 : 111 : 111 111 111 : 111 111 : 111 111
QY      365 PKFWLRVSESE-----VYIAVLQSRKLHAADWAGRARALVGSHTS--WSPA 410
      Db      111 111 111 : 111 : 111 : 111 111 111 : 111 111 : 111 111
QY      385 PQKIHLEDEDEOEGTSEPCCTVLGLGMOKNR-----RQKRIRGOMLSIGAVY 436
      Db      111 111 111 : 111 : 111 : 111 111 111 : 111 111 : 111 111
QY      411 SIPKHYQAVGLHLWKEKRRVNLPRVLSPPVAGTACHAYDREHVLRCELSPCYIAPV 470
      Db      111 111 111 : 111 : 111 : 111 111 111 : 111 111 : 111 111

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 19, 2002, 14:49:21 ; Search time 34.08 Seconds

(without alignments)
3411.168 Million cell updates/sec

Title: US-09-768-877-2

Perfect score: 3649
Sequence: 1 MRAGRGAPARELFRDAFP.....SOEMLGOLFQEVSVNAVMT 672

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 segs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: SP:archaea:*
2: SP:bacteria:*
3: SP:fungi:*
4: SP:human:*
5: SP:invertebrate:*
6: SP:mammal:*
7: SP:mhc:*
8: SP:organelle:*
9: SP:phage:*
10: SP:plant:*
11: SP:rodent:*
12: SP:virus:*
13: SP:vertebrate:*
14: SP:unclassified:*
15: SP:virus:*
16: SP:bacteriophage:*
17: SP:archaeal:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3649	100.0	672	4	09HC96
2	3649	100.0	705	4	096J12
3	3649	100.0	720	4	096J12
4	3383.5	92.7	653	6	095LP4
5	2969	81.4	666	11	099J13
6	2969	81.4	666	11	09ESK3
7	2958	81.1	666	11	09ES66
8	2715.5	74.4	517	4	09HC94
9	2701.5	74.0	544	4	09HC95
10	2699	74.0	513	4	09HC93
11	2691	73.7	605	11	09WVFO
12	2336	64.0	444	4	09HC92
13	1091	29.9	264	11	09HC92
14	858.5	23.5	274	4	09HC91
15	657.5	18.0	720	11	09ER56
16	647	17.7	702	13	0918T0

17	644	17.6	715	13	042133	042133 gallus gall
18	639	17.5	705	13	09YRC1	09YRC1 coturnix co
19	631	17.3	650	6	09N047	09N047 sus scrofa
20	628.5	17.2	716	6	09N185	09N185 bos taurus
21	624	17.1	778	5	09N4B1	09N4B1 caenorhabdi
22	617.5	16.9	640	11	091YU0	091YU0 mus musculu
23	617	16.9	714	6	09GLG2	09GLG2 macaca fasc
24	615.5	16.9	640	11	008688	008688 mus musculu
25	614.5	16.8	703	11	091V43	091V43 mus musculu
26	613.5	16.8	815	6	09GLG7	09GLG7 macaca fasc
27	612	16.8	703	11	064698	064698 ratu
28	611.5	16.8	790	5	09N4B2	09N4B2 caenorhabdi
29	609.5	16.7	709	6	09XSJ3	09XSJ3 oryctolagus
30	609	16.7	737	11	070376	070376 ratu
31	608.5	16.7	757	11	070482	070482 ratu
32	608.5	16.7	785	11	0902F9	0902F9 ratu
33	607	16.6	700	6	09HBB1	09HBB1 homo sapien
34	607	16.6	700	6	09GLG1	09GLG1 macaca fasc
35	606.5	16.6	709	11	008702	008702 ratu
36	604.5	16.6	704	13	09T8G2	09T8G2 brachydanio
37	597	16.4	822	6	09T8H9	09T8H9 bos taurus
38	594.5	16.3	634	4	000263	000263 mus musculu
39	594.5	16.3	709	11	088977	088977 mus musculu
40	593.5	16.3	821	6	046596	046596 sus scrofa
41	592	16.2	502	11	09ER55	09ER55 mus musculu
42	589.5	16.2	709	6	09XSJ1	09XSJ1 bos taurus
43	589	16.1	822	6	09T8H8	09T8H8 ovis aries
44	585	16.0	565	5	09V806	09V806 drosophila
45	583.5	16.0	709	6	09XSJ2	09XSJ2 sus scrofa

ALIGNMENTS

RESULT 1
ID 09HC96 PRELIMINARY; PRT: 672 AA.
AC 09HC96;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE TYPE A CALPAIN-LIKE PROTEASE (CALPAIN 10).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed-11017071;
RA Horikawa Y., Oda N., Cox N.J., Li X., Orho-Melander M., Hara M.,
RA Hinojosa Y., Lindner T.H., Mashima H., Schwarz P., del Bosque Plata L.,
RA Horikawa Y., Oda Y., Yoshinouchi I., Collila S., Polonsky K.S., Wei S.,
RA Concannon P., Iwasaki N., Schulze J., Baier L.J., Bogardus C.,
RA Groop L., Boerwinkle E., Hants C.L., Bell G.I.,
RT Genetic variation in the gene encoding calpain-10 is associated with
RT type 2 diabetes mellitus.*
RL Nat. Genet. 26:163-175(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX PubMed-11017071;
RA Strassberg R.;
RA Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF089088; AAG17966.1; -;
DR EMBL: BC004260; AA04260.1; -;
DR HSSP: Q07009; IDFO;
DR MEROPS: C02.018; -;
DR InterPro: IPR001300; Peptidase_C2.
DR InterPro: IPR000169; Thiolprol_act_sile.
DR Pfam: PF01067; Calpain_III; 2.
DR Pfam: PF00648; Peptidase_C2; 1.
DR PRINTS: PR00704; CALPAIN;
DR SMART: SM00230; Cyspc; 1.
DR PROSITE: PS00139; THIOL_PROTEASE_CYS; UNKNOWN_1.

KW Protease.
SQ SEQUENCE 672 AA; 74938 MW; 74A48D879E997BE1 CRC64;

Query Match 100.0%; Score 3649; DB 4; Length 672;
Best Local Similarity 100.0%; Pred. No. 8e-298;
Matches 672; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRAGRGATPARELFRDAAPFAADSSLCFCDLSTPLAQFREDITWRPQIEICATPRLFPDDP 60
DB 1 MRAGRGATPARELFRDAAPFAADSSLCFCDLSTPLAQFREDITWRPQIEICATPRLFPDDP 60

QY 61 REGQVKGGLGDCWFCACAAALQKSRHLLQVIPPQOPSWADQYRGSTFCRIWQFGRWV 120
DB 61 REGQVKGGLGDCWFCACAAALQKSRHLLQVIPPQOPSWADQYRGSTFCRIWQFGRWV 120

QY 121 EVTTDDRLPCLAGRLCFRCQREDVFWLPLEKVIYAKVHGSYEHLMWAGQVADALVDLTGG 180
DB 121 EVTTDDRLPCLAGRLCFRCQREDVFWLPLEKVIYAKVHGSYEHLMWAGQVADALVDLTGG 180

QY 181 LAERNWLVAGSGGQODRPGWEHRTCRQLLHLKDQCLISCCVLSPPRAGARELGEHAF 240
DB 181 LAERNWLVAGSGGQODRPGWEHRTCRQLLHLKDQCLISCCVLSPPRAGARELGEHAF 240

QY 241 IVSDRLQOAGQCILLRIQNPWGRRCWQGLWREGGEGWSQVDAVAASELLSQLOE 300
DB 241 IVSDRLQOAGQCILLRIQNPWGRRCWQGLWREGGEGWSQVDAVAASELLSQLOE 300

QY 301 FWVEEEFLREFDELTVGYPTVTEAGHLQSLYTERLLCHTRALPGAWYKVGOSAGGCRNSG 360
DB 301 FWVEEEFLREFDELTVGYPTVTEAGHLQSLYTERLLCHTRALPGAWYKVGOSAGGCRNSG 360

QY 361 FPSNPFWLRVSEPVSEYIYAVLQSRHLHAADWAGRALYGDSTSWSPASIPGKHQYQAV 420
DB 361 FPSNPFWLRVSEPVSEYIYAVLQSRHLHAADWAGRALYGDSTSWSPASIPGKHQYQAV 420

QY 421 GLHLWKEKRRVNLPRVLSMPVAGTACHAYDREVLHRCLESPGYILAVPSTFLKDPAGE 480
DB 421 GLHLWKEKRRVNLPRVLSMPVAGTACHAYDREVLHRCLESPGYILAVPSTFLKDPAGE 480

QY 481 FLIRVFSTGRVLSAIRAVAKNTTPGAALPAGEWTVQLRGSWRVGTAGSRNFASYPT 540
DB 481 FLIRVFSTGRVLSAIRAVAKNTTPGAALPAGEWTVQLRGSWRVGTAGSRNFASYPT 540

QY 541 NPCFPSPVPEGPGRVRLTHQCHRPSTDETFHPHIFQVPEGGRSQDAPPLLLQLEPL 600
DB 541 NPCFPSPVPEGPGRVRLTHQCHRPSTDETFHPHIFQVPEGGRSQDAPPLLLQLEPL 600

QY 601 LSCVPHRYAQEVSRLCLLPAGTYKVPSTYLPDTEGAFVTIATRIDRPSIHSEMLGQF 660
DB 601 LSCVPHRYAQEVSRLCLLPAGTYKVPSTYLPDTEGAFVTIATRIDRPSIHSEMLGQF 660

QY 661 LQEVSYMAVYMK 672
DB 661 LQEVSYMAVYMK 672

RESULT 2
Q96J12 PRELIMINARY; PRT: 705 AA.
ID Q96J12
AC Q96J12
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE KIAA1845 PROTEIN (FRAGMENT).
GN KIAA1845
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;

MDLINE=21245130; PubMed=11347906;
RA Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;
RT *Prediction of the coding sequences of unidentified human genes. XX.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.*;
RL DNA Res. 8:95-95(2001).
DR EMBL; AB058748; BAB47474.1; -;
FT NON_TER
SQ SEQUENCE 705 AA; 78287 MW; 3B3BA9824FDF0F53F CRC64;

Query Match 100.0%; Score 3649; DB 4; Length 705;
Best Local Similarity 100.0%; Pred. No. 8.6e-298;
Matches 672; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRAGRGATPARELFRDAAPFAADSSLCFCDLSTPLAQFREDITWRPQIEICATPRLFPDDP 60
DB 34 MRAGRGATPARELFRDAAPFAADSSLCFCDLSTPLAQFREDITWRPQIEICATPRLFPDDP 93

QY 61 REGQVKGGLGDCWFCACAAALQKSRHLLQVIPPQOPSWADQYRGSTFCRIWQFGRWV 120
DB 94 REGQVKGGLGDCWFCACAAALQKSRHLLQVIPPQOPSWADQYRGSTFCRIWQFGRWV 153

QY 121 EVTTDDRLPCLAGRLCFRCQREDVFWLPLEKVIYAKVHGSYEHLMWAGQVADALVDLTGG 180
DB 154 EVTTDDRLPCLAGRLCFRCQREDVFWLPLEKVIYAKVHGSYEHLMWAGQVADALVDLTGG 213

QY 181 LAERNWLVAGSGGQODRPGWEHRTCRQLLHLKDQCLISCCVLSPPRAGARELGEHAF 240
DB 214 LAERNWLVAGSGGQODRPGWEHRTCRQLLHLKDQCLISCCVLSPPRAGARELGEHAF 273

QY 241 IVSDRLQOAGQCILLRIQNPWGRRCWQGLWREGGEGWSQVDAVAASELLSQLOE 300
DB 274 IVSDRLQOAGQCILLRIQNPWGRRCWQGLWREGGEGWSQVDAVAASELLSQLOE 333

QY 301 FWVEEEFLREFDELTVGYPTVTEAGHLQSLYTERLLCHTRALPGAWYKVGOSAGGCRNSG 360
DB 334 FWVEEEFLREFDELTVGYPTVTEAGHLQSLYTERLLCHTRALPGAWYKVGOSAGGCRNSG 393

QY 361 FPSNPFWLRVSEPVSEYIYAVLQSRHLHAADWAGRALYGDSTSWSPASIPGKHQYQAV 420
DB 394 FPSNPFWLRVSEPVSEYIYAVLQSRHLHAADWAGRALYGDSTSWSPASIPGKHQYQAV 453

QY 421 GLHLWKEKRRVNLPRVLSMPVAGTACHAYDREVLHRCLESPGYILAVPSTFLKDPAGE 480
DB 454 GLHLWKEKRRVNLPRVLSMPVAGTACHAYDREVLHRCLESPGYILAVPSTFLKDPAGE 513

QY 481 FLIRVFSTGRVLSAIRAVAKNTTPGAALPAGEWTVQLRGSWRVGTAGSRNFASYPT 540
DB 514 FLIRVFSTGRVLSAIRAVAKNTTPGAALPAGEWTVQLRGSWRVGTAGSRNFASYPT 573

QY 541 NPCFPSPVPEGPGRVRLTHQCHRPSTDETFHPHIFQVPEGGRSQDAPPLLLQLEPL 600
DB 574 NPCFPSPVPEGPGRVRLTHQCHRPSTDETFHPHIFQVPEGGRSQDAPPLLLQLEPL 633

QY 601 LSCVPHRYAQEVSRLCLLPAGTYKVPSTYLPDTEGAFVTIATRIDRPSIHSEMLGQF 660
DB 634 LSCVPHRYAQEVSRLCLLPAGTYKVPSTYLPDTEGAFVTIATRIDRPSIHSEMLGQF 693

QY 661 LQEVSYMAVYMK 672
DB 694 LQEVSYMAVYMK 705

RESULT 3
Q96IG4 PRELIMINARY; PRT: 720 AA.
ID Q96IG4
AC Q96IG4;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE UNKNOWN (PROTEIN FOR IMAGE:2967328) (FRAGMENT).
OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC TISSUE-MELANOMA;
 RA Strausberg R.;
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC007553; AA07553.1; -
 FT NON-TER
 SQ SEQUENCE 720 AA; 79586 MW; 25CD98031078EFD6 CRC64;

Query Match 100.0%; Score 3649; DB 4; Length 720;
 Best Local Similarity 100.0%; Pred. No. 8.8e-298;
 Matches 672; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRAGGATPARELPDPAFPAADSSLFCDLSTPLAQFREDITWRPPOETCATPRLFPDDP 60
 DB 49 MRAGGATPARELPDPAFPAADSSLFCDLSTPLAQFREDITWRPPOETCATPRLFPDDP 108
 QY 61 REGGVKGLDGCWFLCAGALOKSRHLLDQVTPGQPSMADQEQRSFTCRIMQFGMY 120
 DB 109 REGGVKGLDGCWFLCAGALOKSRHLLDQVTPGQPSMADQEQRSFTCRIMQFGMY 168
 QY 121 EVTTDDRLPLCLAGRLCFRCORDEVFMLPLEKYAKYHGSYEHLMAQVADALVDLTGG 180
 DB 169 EVTTDDRLPLCLAGRLCFRCORDEVFMLPLEKYAKYHGSYEHLMAQVADALVDLTGG 228
 QY 181 LAERNMKGAVSGSGQDPRGMEHRTCRQLHLKDOCLISCVLSPRAGARELGEFNAF 240
 DB 229 LAERNMKGAVSGSGQDPRGMEHRTCRQLHLKDOCLISCVLSPRAGARELGEFNAF 288
 QY 241 IVSDLRELQAGOCILLRIQNPWGRRCWGLMREGGEGMSQVDAVAASELLSOLQGE 300
 DB 289 IVSDLRELQAGOCILLRIQNPWGRRCWGLMREGGEGMSQVDAVAASELLSOLQGE 348
 QY 301 FWVEEERFLREFDELTVGYPTTEAGHLOSLYTERILCHTRALPGAMVWGQAGGCRNNG 360
 DB 349 FWVEEERFLREFDELTVGYPTTEAGHLOSLYTERILCHTRALPGAMVWGQAGGCRNNG 408
 QY 361 FFSNPKFRLVSESEYIYIAVLOSRRLAADMAGRARALVDSHTSMSPASIPGKHQAV 420
 DB 409 FFSNPKFRLVSESEYIYIAVLOSRRLAADMAGRARALVDSHTSMSPASIPGKHQAV 468
 QY 421 GLHLMKVEKRRVNLPRVLSMPVAGTACHADREHNLCELSPGYTLAVPSTFLKDAGE 480
 DB 469 GLHLMKVEKRRVNLPRVLSMPVAGTACHADREHNLCELSPGYTLAVPSTFLKDAGE 528
 QY 481 FLRVFSTGRVSLSAITAVAKNTTPGALPAGENGTVOLRGSWVGOTAGSRRNFASPT 540
 DB 529 FLRVFSTGRVSLSAITAVAKNTTPGALPAGENGTVOLRGSWVGOTAGSRRNFASPT 588
 QY 541 NCPFFSVPGPGRCVRIITHQCRPSDMEFHFHIGFHIIFQVPGSGSODAPRLLOEPL 600
 DB 589 NCPFFSVPGPGRCVRIITHQCRPSDMEFHFHIGFHIIFQVPGSGSODAPRLLOEPL 648
 QY 601 LSCVPHRYADEVSRILCLLPAGTYKVPVSTYLPDEGAFVTYIATRIDRPSIHSEMLGOF 660
 DB 649 LSCVPHRYADEVSRILCLLPAGTYKVPVSTYLPDEGAFVTYIATRIDRPSIHSEMLGOF 708
 QY 661 LOEYSVAWAKKT 672
 DB 709 LOEYSVAWAKKT 720

RESULT 4
 ID 095LP4 PRELIMINARY; PRT; 653 AA.
 AC 095LP4;
 DT 01-DEC-2001 (T-EMBLrel. 19, Created)
 DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)

DE HYPOTHETICAL 73 0 KDA PROTEIN.
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecinae; Macaca.
 NCBI_TaxID=9541;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC TISSUE-TESTIS;
 RA Hashimoto K., Osada N., Hida M., Kusuda J., Tanuma R., Hirai M.,
 RA Terao K., Sugano S.;
 RT "Isolation of novel full-length cDNA clones from macaque testis cDNA
 RT libraries";
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB02744; BAB69713.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 653 AA; 72397 MW; 47812497BB315971 CRC64;

Query Match 92.7%; Score 3383.5; DB 6; Length 653;
 Best Local Similarity 95.5%; Pred. No. 1.5e-275;
 Matches 622; Conservative 12; Mismatches 16; Indels 1; Gaps 1;

QY 1 MRAGGATPARELPDPAFPAADSSLFCDLSTPLAQFREDITWRPPOETCATPRLFPDDP 60
 DB 1 MRAGGATPARELPDPAFPAADSSLFCDLSTPLAQFREDITWRPPOETCATPRLFPDDP 60
 QY 61 REGGVKGLDGCWFLCAGALOKSRHLLDQVTPGQPSMADQEQRSFTCRIMQFGMY 120
 DB 61 REGGVKGLDGCWFLCAGALOKSRHLLDQVTPGQPSMADQEQRSFTCRIMQFGMY 120
 QY 121 EVTTDDRLPLCLAGRLCFRCORDEVFMLPLEKYAKYHGSYEHLMAQVADALVDLTGG 180
 DB 121 EVTTDDRLPLCLAGRLCFRCORDEVFMLPLEKYAKYHGSYEHLMAQVADALVDLTGG 180
 QY 181 LAERNMKGAVSGSGQDPRGMEHRTCRQLHLKDOCLISCVLSPRAGARELGEFNAF 240
 DB 181 LAERNMKGAVSGSGQDPRGMEHRTCRQLHLKDOCLISCVLSPRAGARELGEFNAF 240
 QY 241 IVSDLRELQAGOCILLRIQNPWGRRCWGLMREGGEGMSQVDAVAASELLSOLQGE 300
 DB 241 IVSDLRELQAGOCILLRIQNPWGRRCWGLMREGGEGMSQVDAVAASELLSOLQGE 300
 QY 301 FWVEEERFLREFDELTVGYPTTEAGHLOSLYTERILCHTRALPGAMVWGQAGGCRNNG 360
 DB 301 FWVEEERFLREFDELTVGYPTTEAGHLOSLYTERILCHTRALPGAMVWGQAGGCRNNG 360
 QY 361 FFSNPKFRLVSESEYIYIAVLOSRRLAADMAGRARALVDSHTSMSPASIPGKHQAV 420
 DB 361 FFSNPKFRLVSESEYIYIAVLOSRRLAADMAGRARALVDSHTSMSPASIPGKHQAV 420
 QY 421 GLHLMKVEKRRVNLPRVLSMPVAGTACHADREHNLCELSPGYTLAVPSTFLKDAGE 480
 DB 421 GLHLMKVEKRRVNLPRVLSMPVAGTACHADREHNLCELSPGYTLAVPSTFLKDAGE 480
 QY 481 FLRVFSTGRVSLSAITAVAKNTTPGALPAGENGTVOLRGSWVGOTAGSRRNFASPT 540
 DB 481 FLRVFSTGRVSLSAITAVAKNTTPGALPAGENGTVOLRGSWVGOTAGSRRNFASPT 540
 QY 541 NCPFFSVPGPGRCVRIITHQCRPSDMEFHFHIGFHIIFQVPGSGSODAPRLLOEPL 600
 DB 541 NCPFFSVPGPGRCVRIITHQCRPSDMEFHFHIGFHIIFQVPGSGSODAPRLLOEPL 600
 QY 601 LSCVPHRYADEVSRILCLLPAGTYKVPVSTYLPDEGAFVTYIATRIDRPS 650
 DB 601 LSCVPHRYADEVSRILCLLPAGTYKVPVSTYLPDEGAFVTYIATRIDRPS 651

RESULT 5
 ID 099J13 PRELIMINARY; PRT; 666 AA.
 AC 099J13;
 DT 01-JUN-2001 (T-EMBLrel. 17, Created)

01-JUN-2001 (Tremblrel. 17, Last sequence update)
01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE CALPAIN 10.
GN CAPN10.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RA Strausberg R.;
RP Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER, NORMAL. 5 MONTH OLD MALE MOUSE.;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC005681; AAH05681.1; -;
DR EMBL; BC010969; AAH10969.1; -;
DR HSP; Q07009; IDP0.
DR MEROPS; C02.018; -;
DR MGD; MGI:1344392; Capn10.
DR InterPro; IPR001300; Peptidase_C2.
DR InterPro; IPR00169; Thiolprot_act_site.
DR Pfam; PF01067; Calpain_III; 2.
DR Pfam; PF00648; Peptidase_C2; 1.
DR PRINTS; PR00704; CALPAIN.
DR SMART; SM00230; Cyspc; 1.
DR PROSITE; PS00139; THIOI_PROTEASE_CYS; UNKNOWN_1.
SQ SEQUENCE 666 AA; 74596 MW; AOB1B37BE1D2371 CRC64;

Query Match 81.4%; Score 2969; DB 11; Length 666;
Best Local Similarity 81.2%; Pred. No. 9.8e-241;
Matches 545; Conservative 45; Mismatches 75; Indels 6; Gaps 2;

QY 1 MRAGGATPARELFDRAAFPAADSLFCDLSTPLAQFREDITWRPPOICATPLPDDP 60
DB 1 MRVRAETPARELFDRAAFPAADSLFCDLSTPLAQFREDITWRPPOICATPLPDDP 60
QY 61 REGQVKGGLGDCWFLCACAALQKSHLLDQVIPPQSPWADQYRGSTCRINQFGRW 120
DB 61 WEGQVKGGLGDCWFLCACAALQKSHLLDQVIPPQSPWADQYRGSTCRINQFGRW 120
QY 121 EVTTDDRLPCLAGLRCFSRCQREDVFWLPLEKVAHVHAGVADALVDTGG 180
DB 121 EVTTDDRLPCLAGLRCFSRCQREDVFWLPLEKVAHVHAGVADALVDTGG 180
QY 181 LAERNLKVAGSGGQDQRPGRWEHRTCRQLHLKDCCLISCCVLSPRAGARELGEHAF 240
DB 181 LAERNLKVAGSGGQDQRPGRWEHRTCRQLHLKDCCLISCCVLSPRAGARELGEHAF 240
QY 541 NPCPFPSYPEGPGRCVIRITLHQHCRPSDTEFPHIGFHVQVPEGGRSQDAPLLQLQEP 600

DB 535 NPCLPFSYPEGAGPRYIRITLQOHCRLSDSQLHPGFHVQVADGENQDACSLILQEP 594
QY 601 LSCVPHRYAQEVSRCLLPAGTYKVVPSYLPDTEGATVTIATRIDRPSHSQEMQLQF 660
DB 595 LSCVPHRYAQEVSRCLLPAGTYKVVPSYLPDTEGATVTIATRIDRPSHSQEMQLQF 654
QY 661 LQEVSMVAVMK 671
DB 655 LQEVSMVAVMK 665

RESULT 6
Q9ESK3
ID Q9ESK3 PRELIMINARY; PRT; 666 AA.
AC Q9ESK3
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE CALPAIN-LIKE PROTEASE.
GN CAPN10.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20472315; PubMed=11017071;
RA Horikawa Y., Oda N., Cox N.J., Li X., Orho-Melander M., Hara M.,
RA Hinokio Y., Lindner T.H., Mashima H., Schwarz P.E.H.,
RA del Bosque-Platawa L., Horikawa Y., Oda Y., Yoshiuchi I., Collilla S.,
RA Polonsky K.S., Wei S., Concannon P., Iwasaki N., Schulze J.,
RA Baier L.J., Bogardus C., Groop L., Boerwinkle E., Hanis C.L.,
RA Bell G.I.;
RT *Genetic variation in the calpain 10 gene (CAPN10) is associated with
type 2 diabetes.*;
RL Nat. Genet. 26:163-175(2000).
DR EMBL; AF089089; AAG17967.1; -;
DR HSP; Q07009; IDP0.
DR MEROPS; C02.018; -;
DR MGD; MGI:1344392; Capn10.
DR InterPro; IPR001300; Peptidase_C2.
DR InterPro; IPR00169; Thiolprot_act_site.
DR Pfam; PF01067; Calpain_III; 2.
DR Pfam; PF00648; Peptidase_C2; 1.
DR PRINTS; PR00704; CALPAIN.
DR SMART; SM00230; Cyspc; 1.
DR PROSITE; PS00139; THIOI_PROTEASE_CYS; UNKNOWN_1.
KW Protease.
SQ SEQUENCE 666 AA; 74656 MW; BCC8D6778DA74671 CRC64;

Query Match 81.2%; Score 2963; DB 11; Length 666;
Best Local Similarity 81.1%; Pred. No. 3.1e-240;
Matches 544; Conservative 45; Mismatches 76; Indels 6; Gaps 2;

QY 1 MRAGGATPARELFDRAAFPAADSLFCDLSTPLAQFREDITWRPPOICATPLPDDP 60
DB 1 MRVRAETPARELFDRAAFPAADSLFCDLSTPLAQFREDITWRPPOICATPLPDDP 60
QY 61 REGQVKGGLGDCWFLCACAALQKSHLLDQVIPPQSPWADQYRGSTCRINQFGRW 120
DB 61 WEGQVKGGLGDCWFLCACAALQKSHLLDQVIPPQSPWADQYRGSTCRINQFGRW 120
QY 121 EVTTDDRLPCLAGLRCFSRCQREDVFWLPLEKVAHVHAGVADALVDTGG 180
DB 121 EVTTDDRLPCLAGLRCFSRCQREDVFWLPLEKVAHVHAGVADALVDTGG 180
QY 181 LAERNLKVAGSGGQDQRPGRWEHRTCRQLHLKDCCLISCCVLSPRAGARELGEHAF 240
DB 181 LAERNLKVAGSGGQDQRPGRWEHRTCRQLHLKDCCLISCCVLSPRAGARELGEHAF 240
QY 241 IVSDRLQOAGQACIILLRIQNPWGRRCWQGLWREGGEGSWQDAVAASELLSQLEGE 300

Page 5

Db	1	MNAVARETBARELFDIAFAFPA	SSLSFLNPLSLAQFREDITMR	RRPODICTAPOLFDPNP	60	
Qy	61	REGOVAKOGLIDCNCFLCACAL	LOKSRHLLDOVTPPGQSMADQ	EYRSFGTCRIWQGRMV	120	
Db	61	WEGVOXGQGLIGCWFELCACAL	LOKSRHLLDOVFPFGQSMDS	EYQFGFTCRIMQGHME	120	
Qy	121	EYTTDRLREPLCGLGRFCSCF	SCQREDEYFWLPLEKVVAKY	HGSEYEHMAQVADALVDLGS	180	
Db	121	EYTTIDRLPLCGLGRFCSCF	SCQREDEYFWLPLEKAVAKY	HGSEYEHMAQVADALVDLGS	180	
Qy	181	LAERNLKGAVASGGQODRPG	MEHRTCRQLLHLKDQCLISC	CVLSPRAGAREIGEFAF	240	
Db	181	LAERMSLKDIRASGGQODRPS	CGSEHRRACQOLLRLKDQCL	LSCVLSPRAGAREIGEFAF	240	
Qy	241	IYSDIRELOQAGQOCLILRL	IONPMGRNCQGLMRBEGSG	SOYDAVASLSQLDQGE	300	
Db	241	IISDIRELSQGGQILLLRL	HPMGRNCQGLMRBEGSGNO	YVEPAKESILLAQDGE	300	
Qy	301	FVWEDEEFLREDELTGVGY	PVTEAGHLOSILYTERLIC	TRTALPGAMWKGOSAGCRRNSG	360	
Db	301	FVWEDEEFLREDELTIGYPT	EACHLOSILTEKYLCTRAL	PGAWYTGOSAGCRRNSC	360	
Qy	361	FPSNPKFMLRVSESEVYIA	VLQSLRLAADMAGARALV	GDSTHTSPASIPCKHYQAV	420	
Db	361	FPCNPKFMLRLPESEVCVA	VLQRRRL--VGOTRALG--	ASAPAVNLPEKQYQAV	414	
Qy	421	GLHLMKYKRRVNLPRVLS	SNPPVACTACHAYDREYHL	CELSPEYTLAVPSTFLKADPE	480	
Db	415	GLHLMKYKRRKISLPRVLS	APPVAGTACHAADREIHL	CELSPEYTLAVPSTFLKADPGQ	474	
Qy	481	FLIARFSTGRVLSAIRAVAK	NTPGALPAGENGTYOLRSG	SMVGTAGSRRNFASYPT	540	
Db	475	FLIARFSTGRKISLSAVRL	ATGASPGALPAGENETYO	LQGSMAAGOTAGSRRNFASYPC	530	
Qy	541	NCFPFSVPBEGGPRCVRI	TLHOHCRRPSDTEHNF	IGFHIOVPEGRSGRSDAP	PLLLQDEPL	600
Db	535	NCFPFSVPBEGAGPRYIR	ITLQOHCRLSDSQLNHP	IGFHFVFPVADGEEKQDACS	ILLQDEPL	594
Qy	601	LSCVHRARAQVSRCLLPAC	TYKAVPSTYLPDTEGATY	TIAIRIDRPSIHSQEMIGQF	660	
Db	595	LSCVHCACQVSRCLLSA	GNRYLPSTYLPDTEGFTY	TIAIRIDROSISQEMIGOL	654	
Qy	661	LOEVSMAVMK	671			
Db	655	LOEVSFAVMK	665			
RESULT 8						
Q9HC94						
ID	Q9HC94	PRELIMINARY:		PRT:	517 AA.	
AC	Q9HC94:					
DT	01-MAR-2001 (TREMBLrel, 16, Created)					
DT	01-MAR-2001 (TREMBLrel, 16, Last sequence update)					
DE	01-DEC-2001 (TREMBLrel, 19, Last annotation update)					
DE	CALPAIN-LIKE PROTEASE CAPRI0C.					
OS	Homo sapiens (Human).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Plimates; Catarrhini; Homnidae; Homo.					
OX	NCBI_Taxid=9606;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
EX	MEDLINE=20472315; PubMed=11017071;					
RA	Horiikawa Y., Oda N., Cox N.J., Li X., Orho-Melander M., Hara M.,					
RA	Hlnokio Y., Lindner T.H., Mashima H., Schwarz P.E.,					
RA	del Bosque-Plata L., Horiikawa Y. Oda Y., Yoshinchi I., Colilla S.					

RA Baier L.J., Bogardus C., Groop L., Boerwinkle E., Hais C.L.,
RA Bell G.I.;
RT "Genetic variation in the gene encoding calpain-10 is associated with
RT type 2 diabetes mellitus";
RL Nat. Genet. 26:163-175(2000).
DR EMBL; AF089091; AAC17969.1; -.
OR HSP; 007009; IDFO.

```

DR MEROPS: C02.018;
DR InterPro: IPR001300; Peptidase_C2.
DR InterPro: IPR000169; Thiolprot_act_site.
DR Pfam: PF00648; Peptidase_C2; 1.
DR PRINTS: PR00704; CALPAIN.
DR SMART: SM00230; CysPC; 1.
DR PROSITE: PS00139; THIOI_PROTEASE_CYS; UNKNOWN_1.
DR KW: Protease.
DR SEQUENCE 517 AA; 57984 MW; 6761E4FAE4992910 CRC64;

Query Match 74.4%; Score 2715.5; DB 4; Length 517;
Best Local Similarity 76.8%; Pred. No. 1.3e-219;
Matches 516; Conservative 1; Mismatches 0; Indels 155; Gaps 1;

QY 1 MRAGRGATPARELFRDAAFPAADSSLSFCDLSTPLAQFREDITWRRPOEICATPLRFDDP 60
DB 1 MRAGRGATPARELFRDAAFPAADSSLSFCDLSTPLAQFREDITWRRPOEICATPLRFDDP 60
QY 61 REGQVKGGLGDCWFLCACAAQKSRHLLDQVIPPQGPSWADQBYRGSFTCRIQWGRWV 120
DB 61 REGQVKGGLGDCWFLCACAAQKSRHLLDQVIPPQGPSWADQBYRGSFTCRIQWGRWV 120
QY 121 EVTTDDRLPCLAGRLCFSCOREDFVFWLPILLEKYAKVHGSYEHLMAGQVADALVDLTGG 180
DB 121 EVTTDDRLPCLAGRLCFSCOREDFVFWLPILLEKYAKVHGSYEHLMAGQVADALVDLTGG 180
QY 181 LAERNWNLKGVAGSGGQODRGRWEHRTCRQLLHLKDCLISCCVLSPPRAGARELGEHF 240
DB 181 LAERNWNLKGVAGSGGQODRGRWEHRTCRQLLHLKDCLISCCVLSPPRAGARELGEHF 240
QY 241 IVSDRLQOAGQOCILLRIQNPWGRRCWGLWREGGEGSQVDAAVASSELLSQOEGE 300
DB 241 IVSDRLQOAGQOCILLRIQNPWGRRCWGLWREGGEGSQVDAAVASSELLSQOEGE 300
QY 301 FWVEEEFLREFDELTVGYPVTEAGHLQSLYTERLLCHTRALPCAWYKGOSAGCRRNSG 360
DB 301 FWVEEEFLREFDELTVGYPVTEAGHLQSLYTERLLCHTRALPCAWYKGOSAGCRRNSG 360
QY 361 FPSNPKEWLRVSESEYVIAVQBSRLHAADWAGRALVGDSTSWSPASIPKHKYQAV 420
DB 361 FPSNPKEWLRVSESEYVIAVQBSRLHAADWAGRALVGDSTSWSPASIPKHKYQAV 420
QY 421 GLHLWKYKRRVNLPRVLSMPVPVAGTACHAYDREVHLRCELLSPGYLAVPSTFLKDAPGE 480
DB 421 GLHLWK----- 426
QY 481 FLLEVFSTGRVSLSAIRAVAKNTTPGAALPAGEMTVQLRGSWRVGTAGSRNFASYPT 540
DB 427----- 426
QY 541 NPCFPFVSPGPRCVRITLHQHCRPSDTEFFHPIGFHIFQVPEGGESQDAPILLQEP 600
DB 427-----VPEGGESQDAPILLQEP 445
QY 601 LSCVPHRYAQEVSRCLLPAGTYKVVPSYLPDTEGAFVTIATRIDRPSIHQEMLGQF 660
DB 446 LSCVPHRYAQEVSRCLLPAGTYKVVPSYLPDTEGAFVTIATRIDRPSIHQEMLGQF 660
QY 661 LQEVSNMVMNKT 672
DB 506 LQEVSNMVMNKT 517

RESULT 9
Q9HC95
ID Q9HC95 PRELIMINARY; PRT: 544 AA.
AC Q9HC95;
DT 01-MAR-2001 (T-EMBLrel. 16, Created)
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE CALPAIN-LIKE PROTEASE CAPN10B.
OS Homo sapiens (Human).

```


AC 09HC93;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE CALPAIN-LIKE PROTEASE CAPN10D.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_TaxID=9606;
 RX MEDLINE=20472315; PubMed=11017071;
 RA Horikawa Y., Oda N., Cox N.J., Li X., Orho-Melander M., Hara M.,
 RA Hirokio Y., Lindner T.H., Mashima H., Schwarz P.E.,
 RA del Bosque-Plata L., Horikawa Y., Oda Y., Yoshinichi I., Collilla S.,
 RA Polonsky K.S., Wei S., Concannon P., Iwasaki N., Schmeiz J.,
 RA Baier L.J., Bogardus C., Groop L., Boerwinkle E., Hanis C.L.,
 RA Bell G.I.;
 RT "Genetic variation in the gene encoding calpain-10 is associated with
 RT type 2 diabetes mellitus.";
 RL Nat. Genet. 26:163-175(2000).
 DR EMBL: AF089092; AAG17970.1;
 DR HSSP: 007009; 1DFO.
 DR MEROPS: C02.018;
 DR InterPro: IPR001300; Peptidase_C2.
 DR InterPro: IPR000169; Thiolprol_act_site.
 DR Pfam: PF01067; Calpain_III; 1.
 DR Pfam: PF00648; Peptidase_C2; 1.
 DR PRINTS: PR00704; CALPAIN.
 DR SMART: SM00230; Cyspec; 1.
 DR PROSITE: PS00139; THIOI_PROTEASE_CYS; UNKNOWN_1.
 DR KMW Protease.
 SK SEQUENCE 513 AA; 57815 MW; C66DC853F87AEC9C CRC64;

Query Match 74.0%; Score 2699; DB 4; Length 513;
 Best Local Similarity 98.2%; Pred. No. 3.2e-218;
 Matches 496; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 1 MRGRGATPARELFRAAPADSSLFCDSTPLAOFREDITRRROEICATRLPPDP 60
 DB 1 MRGRGATPARELFRAAPADSSLFCDSTPLAOFREDITRRROEICATRLPPDP 60
 QY 61 REGQVQGLLDGDFWFLCACAALOKSRHLDQVIPPQSPWADQEVGSGFTCRIMORGW 120
 DB 61 REGQVQGLLDGDFWFLCACAALOKSRHLDQVIPPQSPWADQEVGSGFTCRIMORGW 120
 QY 121 EYTTDRPLCLAGRLCFSCRCQREDVFWPLLEKVAKVHSGSYEHLMAGVADALVDTGG 180
 DB 121 EYTTDRPLCLAGRLCFSCRCQREDVFWPLLEKVAKVHSGSYEHLMAGVADALVDTGG 180
 QY 121 EYTTDRPLCLAGRLCFSCRCQREDVFWPLLEKVAKVHSGSYEHLMAGVADALVDTGG 180
 DB 121 EYTTDRPLCLAGRLCFSCRCQREDVFWPLLEKVAKVHSGSYEHLMAGVADALVDTGG 180
 QY 181 LAERNMLKGVASGGQDDRGREHRTCRQLHLKQCLISCCVLSPPRAGARELGEFHA 240
 DB 181 LAERNMLKGVASGGQDDRGREHRTCRQLHLKQCLISCCVLSPPRAGARELGEFHA 240
 QY 241 IYSDLEILOGAGOCITLLRIQNPGRRCMOGLMREGSGMSOVDAVAASELLSQOEGE 300
 DB 241 IYSDLEILOGAGOCITLLRIQNPGRRCMOGLMREGSGMSOVDAVAASELLSQOEGE 300
 QY 301 FVVEEEELREFDELTVGYVTEAGHLSLYTERLCHTRALPGAVKQSGAGCCNNNG 360
 DB 301 FVVEEEELREFDELTVGYVTEAGHLSLYTERLCHTRALPGAVKQSGAGCCNNNG 360
 QY 361 FVVEEEELREFDELTVGYVTEAGHLSLYTERLCHTRALPGAVKQSGAGCCNNNG 360
 DB 361 FVVEEEELREFDELTVGYVTEAGHLSLYTERLCHTRALPGAVKQSGAGCCNNNG 360
 QY 421 GLHLMKVERKRNLPVLSMPVAGACHAYDREVLRCESLPGYTLAVPSFLDAPGE 480
 DB 421 GLHLMKVERKRNLPVLSMPVAGACHAYDREVLRCESLPGYTLAVPSFLDAPGE 480
 QY 481 FLIRVFSTGRVSLAIRAVAKMTTP 505
 DB 481 FLIRVFSTGRVSLAIRAVAKMTTP 505

DB 481 FLIRVFSTGRVSLAIRAVAKMTTP 505

RESULT 11
 ID 09WFE0 PRELIMINARY; PRT; 605 AA.
 AC 09WFE0;
 DT 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE CALPAIN-LIKE PROTEASE.
 OS CAPN10 OR CAPN8.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RX MEDLINE=10090;
 RA Braun C., Seifert M., Engel M., Welter C.;
 RA "Identification of a new calpain-like cDNA in mouse lung.";
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF126867; AAD41779.1;
 DR HSSP: 007009; 1DFO.
 DR MEROPS: C02.018;
 DR MGD: MGI:1344392; Capn10.
 DR InterPro: IPR001300; Peptidase_C2.
 DR InterPro: IPR000169; Thiolprol_act_site.
 DR Pfam: PF01067; Calpain_III; 1.
 DR Pfam: PF00648; Peptidase_C2; 1.
 DR PRINTS: PR00704; CALPAIN.
 DR SMART: SM00230; Cyspec; 1.
 DR PROSITE: PS00139; THIOI_PROTEASE_CYS; UNKNOWN_1.
 DR KMW Protease.
 SK SEQUENCE 605 AA; 67788 MW; 85ACEBA43FA0EB12 CRC64;

Query Match 73.7%; Score 2691; DB 11; Length 605;
 Best Local Similarity 80.7%; Pred. No. 1.9e-217;
 Matches 490; Conservative 43; Mismatches 68; Indels 6; Gaps 2;

QY 1 MRGRGATPARELFRAAPADSSLFCDSTPLAOFREDITRRROEICATRLPPDP 60
 DB 1 MRGRGATPARELFRAAPADSSLFCDSTPLAOFREDITRRROEICATRLPPDP 60
 QY 61 REGQVQGLLDGDFWFLCACAALOKSRHLDQVIPPQSPWADQEVGSGFTCRIMORGW 120
 DB 61 REGQVQGLLDGDFWFLCACAALOKSRHLDQVIPPQSPWADQEVGSGFTCRIMORGW 120
 QY 121 EYTTDRPLCLAGRLCFSCRCQREDVFWPLLEKVAKVHSGSYEHLMAGVADALVDTGG 180
 DB 121 EYTTDRPLCLAGRLCFSCRCQREDVFWPLLEKVAKVHSGSYEHLMAGVADALVDTGG 180
 QY 121 EYTTDRPLCLAGRLCFSCRCQREDVFWPLLEKVAKVHSGSYEHLMAGVADALVDTGG 180
 DB 121 EYTTDRPLCLAGRLCFSCRCQREDVFWPLLEKVAKVHSGSYEHLMAGVADALVDTGG 180
 QY 181 LAERNMLKGVASGGQDDRGREHRTCRQLHLKQCLISCCVLSPPRAGARELGEFHA 240
 DB 181 LAERNMLKGVASGGQDDRGREHRTCRQLHLKQCLISCCVLSPPRAGARELGEFHA 240
 QY 241 IYSDLEILOGAGOCITLLRIQNPGRRCMOGLMREGSGMSOVDAVAASELLSQOEGE 300
 DB 241 IYSDLEILOGAGOCITLLRIQNPGRRCMOGLMREGSGMSOVDAVAASELLSQOEGE 300
 QY 301 FVVEEEELREFDELTVGYVTEAGHLSLYTERLCHTRALPGAVKQSGAGCCNNNG 360
 DB 301 FVVEEEELREFDELTVGYVTEAGHLSLYTERLCHTRALPGAVKQSGAGCCNNNG 360
 QY 361 FVVEEEELREFDELTVGYVTEAGHLSLYTERLCHTRALPGAVKQSGAGCCNNNG 360
 DB 361 FVVEEEELREFDELTVGYVTEAGHLSLYTERLCHTRALPGAVKQSGAGCCNNNG 360
 QY 421 GLHLMKVERKRNLPVLSMPVAGACHAYDREVLRCESLPGYTLAVPSFLDAPGE 480
 DB 421 GLHLMKVERKRNLPVLSMPVAGACHAYDREVLRCESLPGYTLAVPSFLDAPGE 480
 QY 481 FLIRVFSTGRVSLAIRAVAKMTTP 505
 DB 481 FLIRVFSTGRVSLAIRAVAKMTTP 505

Page 9

Q9ER56; 01-MAR-2001 (Tremblurel. 16, Created)
 DT 01-MAR-2001 (Tremblurel. 16, last sequence update)
 DT 01-DEC-2001 (Tremblurel. 19, last annotation update)
 DE CALPAIN 12.
 GN CAPN12.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20422668; Pubmed=10964513;
 RA Dear T.N., Meier N.T., Boehm T.:
 RT "Gene structure, chromosomal localization and expression pattern of
 RL Capn12, a new member of the calpain large subunit gene family."; *Genomics* 68:152-160(2000).
 DR EMBL; AJ289241; CAC10066.1; -.
 DR HSSP; Q078009; IDFO.
 DR MEROPS; C02.017; -.
 DR MGD; MGI:1891369; Capn12.
 DR InterPro; IPR002048; EF_hand.
 DR InterPro; IPR001300; Peptidase_C2.
 DR Pfam; PF01067; Calpain_III; 1.
 DR Pfam; PF00648; Peptidase_C2; 1.
 DR PRINTS; PR00704; CALPAIN.
 DR SMART; SM00230; CysPc; 1.
 DR PROSITE; PS00018; EF_Hand; UNKNOWN.1.
 SO SEQUENCE 720 AA; 80588 MW; 37C07BDF0145B531 CRC64;

Query Match	18.0%	Score 657.5;	DB 11;	Length 720;
Best Local Similarity	32.1%	Pred. No. 1.3e-46;		
Matches 170;	Conservative 71;	Mismatches 183;	Indels 105;	Gaps 14

Dd	45	LFRDPCBPADRALGYDKTIGDPSEKAKVGEWKRPHREBAEQFLCEDMSTIDVCGSLN	10
QY	13	LFRDAFPPADSSILFCDLSTPLAQCFREDITWRPOICATRLRPDDPREGQVKGTLG	72
Dd	45	LFRDPCBPADRALGYDKTIGDPSEKAKVGEWKRPHREBAEQFLCEDMSTIDVCGSLN	10
QY	73	CPWLCACALQKSRHLLDQVITPPGQSWADQETKCSFTCRJWQGRNVEYVTDRLPCLA	13
Dd	105	CPWLAASAALSTLYPRLLYVNVPPGQ--GPDQGYAGVEYHFQDLMQGRNVDVYVDKCPVRE	16
QY	133	GRLCFSRCQREDVEFWLPLEEKYAVAHVSYEHLNAGVADALVDLTGSLERNNLTGVAG	19
Dd	163	GKLMFVRSEQRNFEWAPLLEKAYALHDSIYEVMGSGHNNELFVDFTSQVEVLYTLR----	21
QY	193	SGGQODRPRGWEHRTCKROLLH-LKDOCLISCCVLSPPRAGARE--LGEFATVSDLREL	24
Dd	219	-----QMPG-----VPALRRHMLAESLYVGATALSDRGEITRTEGILTKAGHAYSVSTGHH	26
QY	249	QGOAGOCILLRLIONFWGRRCQGLMRBGSGWISQVDAVAASLELSDLGCEFWVEEER	30
Dd	270	-SIGFTKVRLLRLRNWGVKBSGWSQSCRRMDLSEWRDALLVKKEGEEMWMLDQF	32
QY	309	LREPEDELTVGPTVEAGHLSLYTERL-----LCETRLPGAWKQSGASGCRNNS-G	36
Dd	329	LTFEFTNV-----QICSLSPVLGLPSPAGGSMHIIHFGQHWVNGFNGSGQSAEN	37
QY	361	FPSPKPEFWLRYSPSE-----YIANLQKSRH	38
Dd	379	FATNPQFRLLLEPDEEDDDDESGPWGKGACAGPARGKRVKCTYLLSLIQNRP-	43
QY	389	AADMARARALVGDSTHSWSPASIGKHYQVGLH-----LWKYEKRNVNLPVYL	43
Dd	438	-----CLRAKGLT-----YLAVGHVHQIPEELLDLMDSPRSALLPGLL	47
QY	439	SMRPVAGTCAHAYDREYHRLRCLSTGYGLANPSTFLDAPGEFLRYFS	487
Dd	478	R-----ADRSVFCARDVSRCKRLPGHLLVLPVSSASRVGDEDELTLIRFS	522

Thu Sep 19, 15:10:44 2002

us-09-768-877-2.sepi9.rsp

Page 1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 19, 2002, 14:50:01 ; Search time 13.56 seconds
(without alignments)
1918.845 Million cell updates/sec

Title: US-09-768-877-2

Sequence: 1 MRAGRGATPARELFRDAAPF.....SOEMLGQFLQESVWAVKKT 672

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database : SWISSPROT_40.*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	644.5	17.7	702	1	CANB_HUMAN
2	642	17.6	705	1	CANX_CHICK
3	631	17.3	714	1	CANL_PIG
4	629	17.2	700	1	CAN2_MOUSE
5	621.5	17.0	713	1	CAN1_RAT
6	618.5	16.9	714	1	CAN1_HUMAN
7	618	16.9	821	1	CAN3_HUMAN
8	615.5	16.9	713	1	CAN1_MOUSE
9	614	16.8	700	1	CAN2_RAT
10	605.5	16.6	821	1	CAN3_MOUSE
11	604	16.6	700	1	CAN2_HUMAN
12	599.5	16.4	821	1	CAN3_RAT
13	597.5	16.4	700	1	CAN2_CHICK
14	589.5	16.2	810	1	CAN3_CHICK
15	586	16.1	828	1	CAN_DROME
16	579	15.9	653	1	YKR2_CAEEL
17	545	14.9	758	1	CAN_SCHMA
18	512.5	14.0	641	1	CAN6_HUMAN
19	483.5	13.3	641	1	CAN6_MOUSE
20	376.5	10.3	1597	1	SOL_DROME
21	268.5	7.4	422	1	CAN2_RABIT
22	266.5	7.3	374	1	CAN3_PIG
23	114	3.1	528	1	MDLC_PSEPU
24	111	3.0	875	1	NETR_HUMAN
25	110	3.0	3519	1	OL56_STRAT
26	109.5	3.0	116	1	CAN3_BOVIN
27	106	2.9	324	1	CAN2_PIG
28	106	2.9	761	1	NETR_MOUSE
29	105.5	2.9	687	1	TGIC_BOVIN
30	104.5	2.9	673	1	Y552_HUMAN
31	102.5	2.8	686	1	TGIC_MOUSE
32	102.5	2.8	687	1	TGIC_MOUSE
33	102	2.8	573	1	AMH2_HUMAN

ALIGNMENTS

RESULT	1	STANDARD	PRT	702 AA.
ID	CANB_HUMAN			
AC	09UMG6;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DE	Calpain 11 (EC 3.4.22.17) (Calcium-activated neutral proteinase 11)			
DE	(CANP 11).			
GN	CANP11.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=93339389; PubMed=10409436;			
RA	Dear T.N., Moller A., Boehm T.;			
RT	"CAPN11: A calpain with high mRNA levels in testis and located on			
RT	chromosome 6."			
RL	Genomics 59:243-247(1999).			
CC	- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which			
CC	catalyze limited proteolysis of substrates involved in			
CC	cytoskeletal remodelling and signal transduction.			
CC	- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-I-Xaa, Met-I-Xaa or			
CC	Arg-I-Xaa with Leu or Val as the P2 residue.			
CC	- SUBUNIT: Heterodimer of a large (catalytic) and a small			
CC	(regulatory) subunit.			
CC	- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN TESTIS.			
CC	- SIMILARITY: CONTAINS 4 EF-HAND CALCIUM-BINDING DOMAINS.			
CC	- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2; ALSO KNOWN AS THE			
CC	CALPAIN FAMILY OF THIOL PROTEASES.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
CC	-----			
DR	EMBL: AJ242832; CAB52137.1.			
DR	HSSP: P17655; IOKV.			
DR	MIM: 604822.			
DR	InterPro: IPR002048; EF-hand.			
DR	InterPro: IPR001300; Peptidase_C2.			
DR	InterPro: IPR0010169; Thiolprot_act_site.			
DR	Pfam: PF01067; calpain_III.1.			
DR	Pfam: PF00036; ehand.2.			
DR	Pfam: PF00648; peptidase_C2.1.			
DR	PRINTS: PR00704; CALPAIN.			
DR	SMART: SM00230; Cyspc.1.			
DR	SMART: SM00054; Efn.2.			
DR	PROSITE: PS00018; EF_HAND; UNKNOWN 2.			
DR	PROSITE: PS00139; THIOL_PROTEASE_CYS.1.			
DR	PROSITE: PS00639; THIOL_PROTEASE_HIS; FALSE_NEG.			

34	102	2.8	2504	1	FAS_HUMAN	P49327	homo sapien
35	101.5	2.8	302	1	CAN1_RABIT	P06815	oryctolagus
36	100.5	2.8	1463	1	PAZR_BOVIN	P49259	bos taurus
37	99	2.7	517	1	GLPK_MYCTU	O69664	mycobacteri
38	96.5	2.6	429	1	KNIR_DROME	P10734	drosophila
39	95.5	2.6	4303	1	PKD1_HUMAN	P98161	homo sapien
40	95	2.6	457	1	TMS5_HUMAN	O9H353	homo sapien
41	94.5	2.6	808	1	FGFR_MOUSE	O03142	mus sapien
42	94.5	2.6	941	1	GCSP_MYCTU	O50601	mycobacteri
43	94	2.6	748	1	PTIP_SALTY	P37178	salmonella
44	93.5	2.6	689	1	TGIC_CAVCU	P08587	cavia cutle
45	93	2.5	207	1	CAN2_BOVIN	O27971	bos taurus

[illegible]

RESULT	2	CANX_CHECK	STANDARD;	PRT;	705 AA.
ID		CANX_CHECK			
DT	21	P00789;			
DT	21	JUL-1986	(Rel. 01, Created)		
DT	16	OCT-2001	(Rel. 40, Last sequence update)		
DT	16	OCT-2001	(Rel. 40, Last annotation update)		
DE		Calpain, large [catalytic] subunit (EC 3.4.22.17) (Calcium-			
DE		activated neutral proteinase) (CAMP) (Mu/M-type).			
DE		Gallus gallus (Chicken).			
OC		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC		Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;			
OC		Gallus.			
OX		NCBI_TaxID=9031;			
RN		[1]			

[illegible]

SEQUENCE FROM N.A.
MEDLINE=95061606; PubMed=60951110;
Ohno S., Emori Y., Imajob S., Kawasaki H., Kisaragi M., Suzuki K.;
"Evolutionary origin of a calcium-dependent protease by fusion of
genes for a thiol protease and a calcium-binding protein?";
Nature 312:566-570(1984).
[2]
SEQUENCE FROM N.A.
MEDLINE=86082358; PubMed=3000828;
Emori Y., Ohno S., Tobita M., Suzuki K.;
"Gene structure of calcium-dependent protease retains the ancestral
organization of the calcium-binding protein gene.";
FEBS Lett. 194:249-252(1986).
[3]
CALCIUM-BINDING DATA.
MEDLINE=87279982; PubMed=3038855;
Minami Y., Emori Y., Kawasaki H., Suzuki K.;
"E-F hand structure-domain of calcium-activated neutral protease
(CANP) can bind Ca²⁺ ions.";
J. Biochem. 101:889-895(1987).
[4]
CHARACTERIZATION.
MEDLINE=95260862; PubMed=7742367;
Sorimachi H., Tsukahara T., Okada-Ban M., Sugita H., Ishiura S.,
Suzuki K.;
"Identification of a third ubiquitous calpain species -- chicken
muscle expresses four distinct calpains";
Biochim. Biophys. Acta 1261:381-393(1995).
-1- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which
catalyze limited proteolysis of substrates involved in
cytoskeletal remodelling and signal transduction.
-1- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-I-Xaa, Met-I-Xaa or
Arg-I-Xaa with Leu or Val as the P2 residue.
-1- COFACTOR: Binds 3 calcium ions.
-1- ENZYME REGULATION: Activated by micromolar concentrations of
calcium and inhibited by calpastatin.
-1- SUBUNIT: Heterodimer of large (catalytic) and a small (regulatory)
subunit.
-1- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma
membrane upon Ca⁺⁺ binding (By similarity).
-1- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED.
-1- PFM: THE N-TERMINUS IS BLOCKED.
-1- SIMILARITY: Contains 5 EF-hand calcium-binding domains.
-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2; ALSO KNOWN AS THE
CALPAIN FAMILY OF THIOL PROTEASES.
-1- CAUTION: THIS PROTEIN WAS PREVIOUSLY THOUGHT TO BE M-CALPAIN BUT
HAS SINCE BEEN FOUND TO BE AN INTERMEDIATE FORM BETWEEN THE M AND
MU TYPES.

THIS SWISS-PROT entry is copyright. It is produced through a collabora-
tion between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on
use by non-profit institutions as long as its content is in no
modified and this statement is not removed. Usage by and for commer-
cial entities requires a license agreement (See <http://www.isb-sib.ch/announc>
or send an email to license@isb-sib.ch).

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/send an email to license@isb-sib.ch>).

EMBL; X01415; CAA25658.1; -
PIR; A00979; C1CHH.
HSSP; P17655; 1DKV.
MEROPS; C02.003; -
InterPro; IPR002048; EF-hand.
InterPro; IPR001300; Peptidase_C2.
InterPro; IPR000169; Thiolpro1_act_site.
Pfam; PF01067; Calpain_III; 1.
Pfam; PF00036; ehand; 2.
Pfam; PF00648; Peptidase_C2; 1.
PRINTS; PR00704; CALPAIN.
PROSITE; PS00018; EF HAND; 1.
PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
PROSITE; PS00649; THIOL_PROTEASE_ASN; 1.

KM Hydrolase; Thiol protease; Calcium-binding; Multigene family.
 FT DOMAIN 23 213
 FT DOMAIN 214 358 THIOI PROTEASE DOMAIN I.
 FT DOMAIN 359 517 THIOI PROTEASE DOMAIN II.
 FT DOMAIN 518 533 LINKER.
 FT DOMAIN 534 704 DOMAIN IV.
 FT CA_BIND 545 556 EF-HAND 1.
 FT CA_BIND 589 600 EF-HAND 2.
 FT CA_BIND 619 630 EF-HAND 3.
 FT DOMAIN 654 665 ANCESTRAL CALCIUM SITE 4 (POTENTIAL).
 FT DOMAIN 684 695 ANCESTRAL CALCIUM SITE 5 (POTENTIAL).
 FT ACT_SITE 108 108 BY SIMILARITY.
 FT ACT_SITE 265 265 BY SIMILARITY.
 FT ACT_SITE 289 289 BY SIMILARITY.
 SO SEQUENCE 705 AA; ABCDDC5629848AA CRC64;

Query Match 17.6%; Score 642; DB 1; Length 705;
 Best Local Similarity 32.0%; Pred. No. 2.6e-43;
 Matches 162; Conservative 78; Mismatches 188; Indels 78; Gaps 16;

OY 13 LFRDAAPPAOSSL-FCDSLPPLAFREDITWRROEICATRLPPDREGQVKGGLG 71
 DB 48 LFRDQFPAGPPLAFKELG-PYSSKTRGVKRESELVDPOFTYVGATRDICGAGG 106
 OY 72 DCMFLACALAKSRHLLDQVTPGQPSWADQERGSFTCRIMQGRWVETTDRLPCL 131
 DB 107 DCMFLAAGSLNLNELLHRRVPHGOS--FQDYAGIFHFOIMQGEWVYVDDLLPYK 164
 OY 132 AGRCLFSCQKQEDVWPLLEKRVKAVGSEYHLMAGVADALVDLTGGLAEKMLKGYA 191
 DB 165 DGEILFVHSAECTEFWSALLERAKVINGCYESLGGSTTEGEFEDPTGGVAMTDLK-- 221
 OY 192 GSGGQDRGRHNRHETCQLH--LKDCLISCVLSPAGARELGEF--HAF 240
 DB 222 -----RAPRMGHILKALERGLSLCCSIDITSAFDMENATFKLVKGYAY 267
 OY 241 IVSDLRRL--QSOAGQCILLRIQNPGRRCMOGLMREGESQWDAVASELLOE 298
 DB 268 SVTAFKQVNYRGQOQO---LIRIRPMQOVETGMSGSSWMDINDISDRRELQKMD 324
 OY 299 GEFVVEEPELREDELTGVPVTEAGHLQSLYTERL-LCHTRALPGAVKQSGAGCCHN 357
 DB 325 GEFVMSFDFKFEFRLICNLTPDA-----LTKDELNRMTQVEGRTWRSGTAGGCCHN 379
 OY 358 N-SGSPSPKFMRLVSESE-----VYIAVLQSRKLAADWAGRALVGDSTH- 405
 DB 380 NPATFWIMPQFKILLEDDEDDGDEVACSLVALMOKHR-----RRERVGGDMHTI 432
 OY 406 SMSPASIPGKHQVAGVLMK---VEKRRVNLPRVLNMPVAGTACHAYDEVHRLCEL 461
 DB 433 GFAYEVEEAGSGNVHLKKDFLARNOSKRARSEFIML-----REVSNQRL 480
 OY 462 SPGYTLAVPSTFLDAPGEFLRVFS 487
 DB 481 PPEGYIVPSTFEPHKEADFLIRVET 506

RESULT 3
 ID CANI_PIG STANDARD: PRT: 714 AA.
 AC P35750; Q29600; Q9N0M6;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Calpain 1, large [catalytic] subunit (BC 3.4.22.17) (Calcium-activated neutral proteinase) (CANP) (Mu-type) (mucAMP) (Microtubular-calpain).
 GN CAPN1.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
 ON NCBI_TaxID=9623;
 [1]

RP SEQUENCE FROM N.A.
 RA Smith T.P.L., Simmen F.A., Vallet J.A.;
 RT "Rapid Communication: nucleotide sequences of two isoforms of porcine micromolar calcium-activated neutral proteinase 1 (mucalpain) cDNA.";
 RL Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.
 [2]
 RP SEQUENCE OF 326-415 FROM N.A.
 RA Tissue-Small intestine;
 RA Wintere A.K., Fredholm M., Davies W.;
 RL Submitted (AUG-1995) to the EMBL/Genbank/DBJ databases.
 [3]
 RP SEQUENCE OF 528-623 FROM N.A.
 RA Tissue-Skeletal muscle;
 RX MEDLINE-94146155; PubMed-8312396;
 RA Sun W., Ji S.O., Ebert P.J., Bidwell C.A., Hancock D.L.;
 RT "Cloning the partial cDNAs of mu-calpain and m-calpain from porcine skeletal muscle.";
 RT Biochimie 75:931-936(1993).
 CC -1- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which catalyze limited proteolysis of substrates involved in cytoskeletal remodeling and signal transduction.
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-I-Xaa, Met-I-Xaa or Arg-I-Xaa with Ieu or Val as the p2 residue.
 CC -1- COFACTOR: Binds 3 calcium ions (By similarity).
 CC -1- ENZYME REGULATION: Activated by micromolar concentrations of calcium and inhibited by calpastatin.
 CC -1- SUBUNIT: Heterodimer of a large (catalytic) and a small (regulatory) subunit.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma membrane upon Ca++ binding (By similarity).
 CC -1- SIMILARITY: Contains 5 EF-hand calcium-binding domains.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2; ALSO KNOWN AS THE CALPAIN FAMILY OF THIOI PROTEASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AF263610; AAF73444.1;
 DR EMBL: F14611; CA23154.1;
 DR EMBL: 001180; AA65125.1;
 DR HSSP: P17655; IDKV.
 DR MEROPS: C02.001;
 DR InterPro: IPR002048; EF-hand.
 DR InterPro: IPR001300; Peptidase_C2.
 DR InterPro: IPR000169; Thiolprotease.
 DR Pfam: PF01067; Calpain_I; 1.
 DR Pfam: PF00648; Calpain_III; 1.
 DR Pfam: PF00648; Calpain_C2; 1.
 DR PRINTS: PR00704; CALPAIN.
 DR SMART: SM00230; Cysrc; 1.
 DR SMART: SM00054; Eph; 2.
 DR PROSITE: PS00016; EF-HAND; 2.
 DR PROSITE: PS00139; THIOI-PROTEASE_CYS; 1.
 DR PROSITE: PS00639; THIOI-PROTEASE_HIS; FALSE NEG.
 DR PROSITE: PS00640; THIOI-PROTEASE_ASN; FALSE NEG.
 KW Hydrolase; Thiol protease; Calcium-binding; Multigene family.
 FT DOMAIN 30 220 THIOI PROTEASE DOMAIN I.
 FT DOMAIN 221 365 THIOI PROTEASE DOMAIN II.
 FT DOMAIN 366 526 THIOI PROTEASE DOMAIN III.
 FT DOMAIN 527 542 THIOI PROTEASE DOMAIN IV.
 FT DOMAIN 543 713 THIOI PROTEASE DOMAIN V.
 FT CA_BIND 554 565 EF-HAND 1.
 FT CA_BIND 598 609 EF-HAND 2.
 FT CA_BIND 628 639 EF-HAND 3.
 FT CA_BIND 663 674 ANCESTRAL CALCIUM SITE 4 (POTENTIAL).
 FT DOMAIN 693 704 ANCESTRAL CALCIUM SITE 5 (POTENTIAL).
 FT ACT_SITE 115 115 BY SIMILARITY.
 FT ACT_SITE 272 272 BY SIMILARITY.
 FT ACT_SITE 296 296 BY SIMILARITY.

FT	CONFLICT	528	528	V -> I (IN REF. 3).	
FT	CONFLICT	531	531	I -> N (IN REF. 3).	
FT	CONFLICT	541	541	E -> G (IN REF. 3).	
FT	CONFLICT	622	622	S -> A (IN REF. 3).	
SQ	SEQUENCE	714 AA;	81738 MW;	0BB31DE4FC56363A CRC64;	
Query Match 17.3%; Score 631; DB 1; Length 714;					
Best Local Similarity 32.6%; Pred. No. 1.9e-42;					
Matches 167; Conservative 75; Mismatches 182; Indels 88; Gaps 18;					
QY	13	LFDEADPAADSSILFCDLSTPLAQPFEDITWRPQECATPRLFPDDPREGQVKGGLG	71		
DB	55	LFDEADPAADSSILFCDLSTPLAQPFEDITWRPQECATPRLFPDDPREGQVKGGLG	71		
QY	72	DCWFLCACAALOKSRHLLQDQVPPGQSWADQYRGSEFCRIWQFGRWVETDRLPCL	131		
DB	114	DCWFLCACAALOKSRHLLQDQVPPGQSWADQYRGSEFCRIWQFGRWVETDRLPCL	131		
QY	132	AGRLCSRCQREDVFWLPLEKYAKVHGYHLWAGOVADALVDLTGGLAERWNLKQVA	191		
DB	172	DGKLVFVHSAQGNFWSALLERAKAYKNGSYEALSGGSTSEGEDFTGGVTE-W-----	224		
QY	192	GGSGQQDRPCGRWHRCTCQLLH-----LKDQCLISC-----CYLSPRAGA-RELGEFHA	239		
DB	225	-----YELKAPSDLSIILKALERSGLGCSIDISSVLDMEAVTFKLVKGHA	273		
QY	240	FIVDLRELQAGQACILLIRONPGRRCWQGLMRGEGGSOVDAAVASSELLSOLQEG	299		
DB	274	YSVTGAKQNYQ-GQMVNLIRNPNWGEVETWGANSGSEWGVDPYQDQLRWREDOG	332		
QY	300	EFWVEEFLREDELTGVPTVTEAGHLQSLYTERLL-CHTRALPGAWKVGOSAGCRNN	358		
DB	333	EFWVSPRDFLREFTLEICNLTPDA-----LKSQRVNNWNTTLYEGTWRGSTAGSCRN-	386		
QY	359	SGPPS-----NPKFLVASE-----PSEVY-----IAVLQSRHLHADWAGRALVNG	401		
DB	387	-----YPAFTWNPQFKIRLEETDDPDYGGREGSCSVTALMOKH-----RRERF----	435		
QY	402	DSHTSWSPASIPKHYQAVGLHMKVKKRVNLPVLSPMPVAGTACHAYD-----REV	455		
DB	436	-----GRDMETIGFAVYVPELVQPVHLKRDFFLANASRARSEQFINLREV	483		
QY	456	HLRCELSPGYLLAVSTFKLDAPGELLVFS	487		
DB	484	STRFLPPGVEYVVFSTFEPEKNGDFVLRFSS	515		
RESULT 4					
CAN2_MOUSE STANDARD; PRT; 700 AA.					
ID	CAN2_MOUSE	STANDARD;	PRT;	700 AA.	
AC	008529;	035518;	054843;		
DT	01-NOV-1997	(Rel. 35, Created)			
DT	16-OCT-2001	(Rel. 40, Last sequence update)			
DT	01-MAR-2002	(Rel. 41, Last annotation update)			
DE	Calpain 2, large [catalytic] subunit precursor (EC 3.4.22.17).				
DE	(Calcium-activated neutral proteinase) (CANP) (M-type) (M-calpain)				
DE	(Millimolar-calpain) (80 kDa M-calpain subunit) (CALP80).				
GN	CANP2.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-BALB/C;				
EX	MEDLINE=97480729; PubMed=9339374;				
RA	Dear N., Matena K., Vingron M., Boehm T.;				
RT	*A new subfamily of vertebrate calpains lacking a calmodulin-like				
RL	domain: implications for calpain regulation and evolution.*;				
RL	Genomics 45:175-184(1997).				
RN	[2]				
RP	SEQUENCE FROM N.A.				

RC	STRAIN-BALB/C;				
RA	Ozaki Y.;				
RL	Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=CNS;				
RA	Glass J.D., Nash N.R., Dry I., Culver D., Wesselingh S.;				
RT	*Cloning of m-calpain from mouse nervous system.*;				
RL	Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.				
CC	-!- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which				
CC	catalyze limited proteolysis and signal transduction involved in				
CC	cytoskeletal remodeling and signal transduction (By similarity).				
CC	-!- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-I-Xaa, Met-I-Xaa or				
CC	Arg-I-Xaa with Leu or Val as the P2 residue.				
CC	-!- COFACTOR: Binds 3 calcium ions.				
CC	-!- ENZYME REGULATION: Activated by 200-1000 micromolar concentrations				
CC	of calcium and inhibited by calpastatin.				
CC	-!- SUBUNIT: Heterodimer of a large (catalytic) and a small				
CC	(regulatory) subunit.				
CC	-!- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma				
CC	membrane upon Ca++ binding.				
CC	-!- SIMILARITY: CONTAINS 5 EF-HAND CALCIUM-BINDING DOMAINS.				
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2; ALSO KNOWN AS THE				
CC	CALPAIN FAMILY OF THIOL PROTEASES.				

This SWISS-PROT entry is copyright. It is produced through a collaboration					
between the Swiss Institute of Bioinformatics and the EMBL outstation -					
the European Bioinformatics Institute. There are no restrictions on its					
use by non-profit institutions as long as its content is in no way					
modified and this statement is not removed. Usage by and for commercial					
entities requires a license agreement (See http://www.isb-sib.ch/announce/					
or send an email to license@isb-sib.ch).					

EMBL;	Y10139;	CAA71227.1;			
EMBL;	D38117;	BAA22964.1;			
EMBL;	AF015038;	AAB94029.1;			
HSSP;	P17655;	1DKV.			
MEROPS;	C02.002;				
MGD;	MGI:88264;	Capn2.			
InterPro;	IPR002048;	EF-hand.			
InterPro;	IPR001300;	Peptidase_C2.			
InterPro;	IPR000169;	Thiolprot_act_site.			
Pfam;	PF01067;	Calpain.III.1.			
Pfam;	PF00036;	efhand.3.			
Pfam;	PF00648;	Peptidase_C2.1.			
PRINTS;	PR00704;	CALPAIN.			
SMART;	SM00230;	CysPc.1.			
SMART;	SM00054;	Efh.2.			
PROSITE;	PS00018;	EF HAND.2.			
PROSITE;	PS00139;	THIOL PROTEASE_CYS.1.			
PROSITE;	PS00639;	THIOL PROTEASE_HIS; FALSE_NEG.			
PROSITE;	PS00640;	THIOL PROTEASE_ASN; FALSE_NEG.			
KW	Hydrolase; Thiol protease; Calcium-binding; Repeat; Multigene family.				
FT	PROPEP	1 19			
FT	CHAIN	20 700			
FT	DOMAIN	20 210			
FT	DOMAIN	211 355			
FT	DOMAIN	356 514			
FT	DOMAIN	515 529			
FT	DOMAIN	530 700			
FT	DOMAIN	530 700			
FT	CA_BIND	541 552			
FT	CA_BIND	585 596			
FT	CA_BIND	615 626			
FT	DOMAIN	650 661			
FT	DOMAIN	680 691			
FT	ACT_SITE	105 105			
FT	ACT_SITE	262 262			
FT	ACT_SITE	286 286			
FT	CONFLICT	194 194			
FT	CONFLICT	212 212			
FT	CONFLICT	402 402			
SQ	SEQUENCE	700 AA;	79871 MW;	682146B290968316 CRC64;	

Query Match 17.2% Score 629; DB 1; Length 700;
 Best Local Similarity 30.7% Pred. No. 2.7e-42;
 Matches 158; Conservative 78; Mismatches 207; Indels 72; Gaps 12;

QY 10 ARELEDAAPPAADSLCDLSTPLAQRREDITWRROEICATPPLFPDDPREGQKGL 69
 DB 42 AGALFQDPSFPALPSLSQKELGYSSTKRGIEKMRPEICADPOFIIGATRDICGA 101
 QY 70 LGDCWFLCACAALOKSRHLLDQVTPPGPSWADDERGSCFCRIQWGFVETDDSLP 129
 DB 102 LGDCWFLAALAIASLTINEELIARVPPQS--FQENVAGIFHOFQWQGEWEVYDDSLP 159
 QY 130 CLARLFCRQRREDVWPLPLEKVVAVHGSYTEHLMAGVADALVDLTGGLAERWNLKG 189
 DB 160 TKOELLFVHSAEGSEFMSALLEKAYAKINCYEALSGATTEGPEFTGGIAEYELR- 218
 QY 190 VAGSGQODRPRGHEHRTKROLHLKQOCLISCVLSRAGAR-----LGEHAFIVS 243
 DB 219 -----KRPPLFKTIQKA--LEKSLGCSIDITSADSEAVYQKLYKGHAYSVT 267
 QY 244 DLRELQAOACQILLRIQWGRRCWGLMREGSGMSQVDAVAASELSQLEGEFVW 303
 DB 268 GAEEVE--SGSLQKLRIRNPMQGVETGKMNDCPSMNTVDPRANLTEROEDGEFWM 326
 QY 304 EEEFLREFDELTVGYVTEAGHLQSLYTERLCHT-----RALPGAWMGOSAGGCRN 357
 DB 327 SFSDFLHYSRL-----EICNLTPDLTCDSYKKMKLTFKMDGNMRRGSAAGCGRN 376
 QY 358 -NSGFSPNPFMLRVSESE-----VYIAVLRSLHADNAGRAVALGDSHTS 406
 DB 377 YPRTFMNPPQYLKLEDEDEDEDEGERGCTFLVGLQKHN-----RROKRGEDHMTI 429
 QY 407 W-----SPASIPCKHYQAVGLHLMKVEKRRVNLPRVLSMPVAGTACHADREYHLCCL 461
 DB 430 GEGIVPEPELTQNTNHLCKNFFLTFRABERSDPTFNL-----REVLANRKL 477
 QY 462 SPGYLAVPSTFLKADAPGERFLAVFSTGRVLSAI 496
 DB 478 PGEIVLPSTPEPRKDGFCIRVFESEKADYQAV 512

RESULT 5
 CAN1_RAT STANDARD; PRT; 713 AA;
 AC P97571;
 DT 16-OCT-2001 (Rel. 40; Created)
 DT 16-OCT-2001 (Rel. 40; Last sequence update)
 DE Calpain 1, large [catalytic] subunit (EC 3.4.22.17) (Calcium-activated neutral protease) (CAMP) (Mu-type) (mucAMP) (Micromolar-calpain).
 GN CAMP1 OR CLS1.
 OS Rattus norvegicus (Rat).
 OC Eutheria; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_Taxid=10116;
 RN
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97107433; PubMed=8950173;
 RA Sotimachi H., Amano S., Ishiura S., Suzuki K.;
 RT Primary sequences of rat mu-calpain large and small subunits are,
 RT respectively, moderately and highly similar to those of human.
 RL Blochm. Biophys. Acta 1309:37-41(1996).
 CC -1- FUNCTION: Calcium regulated non-lysosomal thiol-protease which
 CC catalyze limited proteolysis of substrates involved in
 CC cytoskeletal remodeling and signal transduction.
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-I-Xaa, Met-I-Xaa or
 CC Arg-I-Xaa with Leu or Val as the p2 residue.
 CC -1- COFACTOR: Binds 3 calcium ions (by similarity).
 CC -1- ENZYME REGULATION: Activated by micromolar concentrations of
 CC calcium and inhibited by calpastatin.
 CC -1- SUBUNIT: Heterodimer of a large (catalytic) and a small
 CC (regulatory) subunit.

CC -1- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma
 CC membrane upon Ca++ binding (By similarity).
 CC -1- SIMILARITY: Contains 5 EF-hand calcium-binding domains.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2; ALSO KNOWN AS THE
 CC CALPAIN FAMILY OF THIOL PROTEASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; U53858; AAC53001.1; -
 DR HSSP; P17655; IDKV.
 DR MEROPS; C02.001; -
 DR Interpro; IPR002048; EF-hand.
 DR Interpro; IPR001300; Peptidase_C2.
 DR Interpro; IPR000169; Thiolprol_act_site.
 DR Pfam; PF01067; Calpain_III; 1.
 DR Pfam; PF00036; ehand; 3.
 DR Pfam; PF00648; Peptidase_C2; 1.
 DR PRINTS; PR00704; CALPAIN.
 DR SMART; SM00230; CysPc; 1.
 DR SMART; SM00054; Efr; 2.
 DR PROSITE; PS00018; EF_HAND; 2.
 DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
 DR PROSITE; PS00639; THIOL_PROTEASE_HIS; FALSE NEG.
 DR PROSITE; PS00640; THIOL_PROTEASE_ASP; FALSE NEG.
 KW Calcium-binding; Hydrolase; Thiol protease.
 FT DOMAIN 30 220
 FT 221 365 THIOL PROTEASE DOMAIN I.
 FT DOMAIN 366 525 THIOL PROTEASE DOMAIN II.
 FT DOMAIN 526 541 LINKER.
 FT DOMAIN 542 712 DOMAIN III, C2-LIKE DOMAIN.
 FT CA_BIND 553 564
 FT CA_BIND 597 608
 FT CA_BIND 627 638
 FT DOMAIN 662 673
 FT DOMAIN 692 703 ANCESTRAL CALCIUM SITE 3 (POTENTIAL).
 FT ACT_SITE 115 115 ANCESTRAL CALCIUM SITE 4 (POTENTIAL).
 FT ACT_SITE 272 272 BY SIMILARITY.
 FT ACT_SITE 296 296 BY SIMILARITY.
 SQ SEQUENCE 713 AA; 82118 MW; 6E664600B0FABEBB CRC64;

Query Match 17.0% Score 621.5; DB 1; Length 713;
 Best Local Similarity 31.8% Pred. No. 1.1e-41;
 Matches 161; Conservative 80; Mismatches 188; Indels 77; Gaps 15;

QY 13 LFRDAAPPAADSL-FCDSLTPLAQREDITWRROEICATPPLFPDDPREGQKGLG 71
 DB 55 LFRDDAFPPVSHSLGFEKELG-PNRSKYTGIKMRPELISNPFIYDGAITRDICGALG 113
 QY 72 DCMFLCACAALOKSRHLLDQVTPPGPSWADDERGSCFCRIQWGFVETDDSLP 131
 DB 114 DCMFLAALAIASLTINEELIARVPPQS--FQENVAGIFHOFQWQGEWEVYDDSLP 171
 QY 132 ASRLFCRQRREDVWPLPLEKVVAVHGSYTEHLMAGVADALVDLTGGLAERWNLKGVA 191
 DB 172 DCKLVFVHSAEGSEFMSALLEKAYAKVNSYERALSQCSSEAFEDTGGVTEYD----- 227
 QY 192 GSGGQODRPRGHEHRTKROL--HLKQOCLISCVLSRAGARLEGEF-----HAFIVS 243
 DB 228 -----OKAPS-----DLYOIILKALERSGLACISINISDIRLEAITFRNLRGHAYSVT 277
 QY 244 DLRELQAOACQILLRIQWGRRCWGLMREGSGMSQVDAVAASELSQLEGEFVW 303
 DB 278 DAKQVTVYQGVAVNLIRNPMQGVETGKMNDCPSMNTVDPRANLTEROEDGEFWM 336
 QY 304 EEEFLREFDELTVGYVTEAGHLQSLYTERLCHTALPGAWMGOSAGGCRNNSGSPS 363

Db 337 SFDFIRETKLEICNLPDALKSRLRN-----WNTTYEGTWRRGSTRAGGCRN---YPA 389
 QY 364 ----NPKFWLRVSEFSEV-----YIATLQSRHLHAADWAGRALVGDSTSW 407
 Db 390 TFWNPFQKIRLEVDADDYDSRSGCSFLALAKQHR-----RRERF-----434
 QY 408 SPASIPGHHYQVGLHKLKVEKRRVNLPRVLSMPVPVAGTACHAYD-----REVHLRCEL 461
 Db 435 -----GRDMETIGFANVQVRELGAQPVHLKRDFFLANASQAQSEHFINLREVSRL 488
 QY 462 SPGYVLAVPSTFLKDPAGEFLLRVPS 487
 Db 489 PPGCIYVVPSTFEPNKGDFLLRFFS 514

RESULT 6
 ID CAN1_HUMAN STANDARD; PRT; 714 AA.
 AC P07384;
 DT 01-APR-1988 (Rel. 07, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Calpain 1, large [catalytic] subunit (EC 3.4.22.17) (Calcium-activated
 DE neutral proteinase) (CAMP) (Mu-type) (muCAMP) (Micromolar-calpain).
 GN CAPN1 OR CAPN1L.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86301172; PubMed=3017764;
 RA Aoki K., Imajoh S., Ohno S., Emori Y., Koike M., Kosaki G., Suzuki K.;
 RT "Complete amino acid sequence of the large subunit of the low-Ca2+-
 RT requiring form of human Ca2+-activated neutral protease (muCAMP)
 RT deduced from its cDNA sequence.";
 RL FEBS Lett. 205:313-317(1986).
 RN [2]

SEQUENCE FROM N.A.
 RX MEDLINE=90380278; PubMed=2400579;
 RA Sorimachi H., Ohmi S., Emori Y., Kawasaki H., Saido T.C., Ohno S.,
 RA Minami Y., Suzuki K.;
 RT "A novel member of the calcium-dependent cysteine protease family.";
 RL Biol. Chem. Hoppe-Seyler 371:171-176(1990).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RA Strausberg R.;
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which
 CC catalyze limited proteolysis of substrates involved in
 CC cytoskeletal remodeling and signal transduction.
 CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-I-Xaa, Met-I-Xaa or
 CC Arg-I-Xaa with Leu or Val as the P2 residue.
 CC -!- COFACTOR: Binds 3 calcium ions.
 CC -!- ENZYME REGULATION: Activated by micromolar concentrations of
 CC calcium and inhibited by calpastatin.
 CC -!- SUBUNIT: Heterodimer of a large (catalytic) and a small
 CC (regulatory) subunit.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma
 CC membrane upon Ca++ binding (By similarity).
 CC -!- SIMILARITY: Contains 5 EF-hand calcium-binding domains.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2; ALSO KNOWN AS THE
 CC CALPAIN FAMILY OF THIOL PROTEASES.

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

DR EMBL; X04366; CAA27881.1; --
 DR EMBL; BC008751; AA08751.1; --
 DR PIR; A26213; CIRUH.
 DR PIR; S10591; S10591.
 DR HSP; P17655; IDKV.
 DR MEROPS; C02.001; --
 DR MIN; I14220; --
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR001300; Peptidase_C2.
 DR InterPro; IPR000169; Thiolprot_act_site.
 DR Pfam; PF01067; Calpain_III; 1.
 DR Pfam; PF00036; ehand; 3.
 DR Pfam; PF00648; Peptidase_C2; 1.
 DR PRINTS; PR00704; CALPAIN.
 DR SMART; SM00230; CysPc; 1.
 DR SMART; SM00034; EFh; 2.
 DR PROSITE; PS00018; EF_HAND; 2.
 DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
 DR PROSITE; PS00639; THIOL_PROTEASE_HIS; FALSE_NEG.
 DR PROSITE; PS00640; THIOL_PROTEASE_ASN; FALSE_NEG.
 KW Hydrolase; Thiol protease; Calcium-binding; Multigene family.
 FT DOMAIN 30 220 THIOL PROTEASE DOMAIN I.
 FT DOMAIN 221 365 THIOL PROTEASE DOMAIN II.
 FT DOMAIN 366 526 DOMAIN III, C2-LIKE DOMAIN.
 FT DOMAIN 527 542 LINKER.
 FT DOMAIN 543 713 DOMAIN IV.
 FT CA_BIND 554 565 EF-HAND 1.
 FT CA_BIND 598 609 EF-HAND 2.
 FT CA_BIND 628 639 EF-HAND 3.
 FT DOMAIN 663 674 ANCESTRAL CALCIUM SITE 4 (POTENTIAL).
 FT DOMAIN 693 704 ANCESTRAL CALCIUM SITE 5 (POTENTIAL).
 FT ACT_SITE 115 115 BY SIMILARITY.
 FT ACT_SITE 272 272 BY SIMILARITY.
 FT ACT_SITE 296 296 BY SIMILARITY.
 FT CONFLICT 548 548 K -> N (IN REF. 3).
 SQ SEQUENCE 714 AA; 81889 MW; 1CB6D7C56D063498 CRC64;

Query Match 16.9%; Score 618.5; DB 1; Length 714;
 Best Local Similarity 32.3%; Pred. No. 1.9e-41;
 Matches 166; Conservative 73; Mismatches 198; Indels 77; Gaps 17;
 QY 13 LFRDAAPPAADSSL-FCDLSTPLAQFREDITWRPQECATPRLPDPDPKQVGLG 71
 Db 55 LFRDAEPFPVQSLGKDLG-PNSKTYGIKWRTELLSNPQFIVDQATRTDICOAGL 113
 QY 72 DCWELCAALQKSRHLDDVIPPQPSWADQYRGSTCRWQFGRVVEVTDRLPCL 131
 Db 114 DCWLLAATASLTNDTLHVRVPHQS--FQNGYAGIFHQLQWQGVVVDVLLPIK 171
 QY 132 AGRICFSCQREDVFWLPLEKVKYAKVHSGYHLWAGQVADALVDLTGGLAERNLKGVA 191
 Db 172 DGLVVFVHSAEGNEFWSALEKAYAKVNSYEALSGGSTSEGFEDFTGCVTEWELRRAP 231
 QY 192 GSGGQDRPGRWEHRTCRLLHLKDDCLISC-----CVLSPPA-GARELGFEHAFIVSDL 245
 Db 232 SLDLQ-----IILKALRGSLGCSIDISSVLDMEATFKKLVKGHAYSVTGA 279
 QY 246 RELQAGQOCILLRIQNPWGRRCQGLWREGGEGVSDAASVASELISQEGEFWVEE 305
 Db 280 KQV-NYRGQVYVLLMRNPNCEVEWTGAWSDSSSENWNVDPYERDQLRVKMEDECFWNSP 338
 QY 306 EEFRLREFDELTVGPVTEAGHLQSLYTERLLCHTRALPGAWVKQSGAGGCRNNSGFPSS-- 363
 Db 339 RDMREFRLEICNLTPDA--LKSRTIRK--WNTTLYEGTWRRGSTRAGGCRN---YPATF 391
 QY 364 --NPKFWLRVSESE-----YIATLQSRHLHAADWAGRALVGDSTSW 409
 Db 392 WVNPFQKIRLETDVDDYDGRSGCSFVLALAKQHR-----RRERF-----434
 QY 410 ASIPGHHYQVGLHKLKVEKRRVNLPRV-LSMPPVAGTACHAYD-----REVHLRCELS 462
 Db 435 ----GRDMETIGFAYVPPVELVGQPAVHLKRDFFLANASPARSEQFINLREVSRL 490

Kurnit D.M., Lefranc G., Legum C., Loiselet J., Merlini L.

CC -1- DATABASE: NAME=Leiden Muscular Dystrophy pages;


```

CC      membrane upon Ca++ binding (By similarity).
CC      -1- SIMILARITY: Contains 5 EF-hand calcium-binding domains.
CC      -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2; ALSO KNOWN AS THE
CC      CALPAIN FAMILY OF THIOL PROTEASES.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AF021847; AAB72222.1;
DR      EMBL; AF084459; AAC33134.1;
DR      HSSP; P17655; IDKV.
DR      MEROPS; C02.001;
DR      MGD; MGI:88263; Capn1.
DR      InterPro; IPR002048; EF-hand.
DR      InterPro; IPR001300; Peptidase_C2.
DR      InterPro; IPR00169; Thiolprot_act_site.
DR      Pfam; PF01067; Calpain_III; 1.
DR      Pfam; PF00036; ehfand_3.
DR      Pfam; PF00648; Peptidase_C2; 1.
DR      PRINTS; PR00704; CALPAIN.
DR      SMART; SM00230; Cyspc; 1.
DR      SMART; SM00504; Eph; 2.
DR      PROSITE; PS00018; EF_HAND; 2.
DR      PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
DR      PROSITE; PS00639; THIOL_PROTEASE_HIS; FALSE NEG.
DR      PROSITE; PS00640; THIOL_PROTEASE_ASN; FALSE NEG.
KW      Hydrolyase; Thiol protease; Calcium-binding; Multigene family.
FT      DOMAIN 30 220
FT      DOMAIN 221 365
FT      DOMAIN 366 525
FT      DOMAIN 526 541
FT      DOMAIN 542 712
FT      CA_BIND 553 564
FT      CA_BIND 597 608
FT      CA_BIND 627 638
FT      DOMAIN 662 673
FT      ACT_SITE 692 703
FT      ACT_SITE 115 115
FT      ACT_SITE 272 272
FT      ACT_SITE 296 296
FT      CONFLICT 488 488
FT      CONFLICT 696 696
SQ      SEQUENCE 713 AA; 82106 MW; 3E1E2C695802B864 CRC64;

```

Query Match 16.98; Score 615.5; DB 1; Length 713;
 Best Local Similarity 32.08; Pred. No. 3.3e-41;
 Matches 162; Conservative 80; Mismatches 187; Indels 77; Gaps 16;

```

QY 13 LFEDAPPADSSLT-FCDLSTPLAOFREDITWRROICATPRLPDPDPREQVAGGLG 71
DB 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
QY 55 LFQDAFPVSHSLGFKELG-PHSKITYGIMKRTPELTMSNPOFLVDATRTDICOGLG 113
DB 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
QY 72 DCHFLCACALQKSRHLLDOVTPGQPSMADEYRGSTFCRIMQFGRVETVTDRLCL 131
DB 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
QY 114 DCHMLAALASLTLETTLIRVVPYGQ-SFQD-GYAGIFHPDLMQPGEVNDVVDLLE 171
DB 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
QY 132 AGRICFSCNQEDVFWPLILEKYAKYHGSTEHLMAGQVADALVDLGLAERNLKGVA 191
DB 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
QY 172 DGLVYFHSAGNENFQWALLKAKAKYNGSEYALSGCTSEAFEDFPGVTEWDL- 227
DB 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
QY 192 GSGGQDPRGWEHRTGRLT-HLKQDCIISCCVSLPRAGARLGEF-----HAFIVS 243
DB 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
QY 228 -----QKAPS-----DIQYIIKLKRLGSLGCSINISDIRDLFAIRFKMLVNRCHAYSVT 277
DB 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
QY 244 DLRELQAGQCCILLRIQNPWGRRCWGLMRGEGSGMGOVDAVAASELSLOLGEFMY 303
DB 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
QY 278 GARQVYQY-GQVNLILMRPMEVEEMKGPWMSQSYEMNKNVDYERKDLVKKMDDGEFM 336
DB 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11

```

```

QY 304 EEEFIREDELTYGYPTVEAGHLSLTERLCHTRALPGAWKGSAGCCRNNGSPS 363
DB 337 SFROPIREFTLEICNLTPDALKSRITLNN-----NATFEYEGTWGRSGTAGCCRN---YPA 389
QY 364 ---NPKRWLVSEPSSEV-----YIAVQNSRLHAADWAGRALVQDSRTSW 407
DB 390 TFVWNPQFKIRLEVDADVDNRESGSEFLALMOKHR-----RRERF----- 434
QY 408 SPASIPKHYQAVCLHMKYKRRVNI-PRVLSMPVAGTACHAYD-----REYHLRCEL 461
DB 435 -----GRDMEITGFAYQVPRELAGQPVHLKRDFPLANMSRAQSEHFINREVSNI 488
QY 462 SPGYLAVPSTFLKADAGEFLIRVES 487
DB 489 PGGEYIVPSTFEPNKGDFLIRFES 514

```

RESULT 9

```

ID CAN2_RAT STANDARD; PRT; 700 AA.
AC 007009;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Calpain 2, large (catalytic) subunit precursor (EC 3.4.22.17)
DE (Calcium-activated neutral proteinase) (CAMP) (M-type) (M-calpain)
DE (Millimolar-calpain).
GN CAPN2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-94032492; PubMed-8218419;
RA Deluca C.I., Davies P.L., Samis P.L., Elce J.S.;
RT "Molecular cloning and bacterial expression of cDNA for rat calpain
RL II 80 kDa subunit.";
RL Biochim. Biophys. Acta 1216:81-93(1993).
RN [2]
RP PARTIAL SEQUENCE.
RX MEDLINE-21240297; PubMed-11342050;
RA Moldoveanu T., Hosfield C.M., Jia Z., Elce J.S., Davies P.L.;
RT "Ca(2+)-induced structural changes in rat m-calpain revealed by
RL partial proteolysis.";
RL Biochim. Biophys. Acta 1545:245-254(2001).
RN [3]
RP MOTAGENESIS OF LYS-230; LYS-234 AND GLU-504.
RX MEDLINE-21269273; PubMed-11102442;
RA Hosfield C.M., Moldoveanu T., Davies P.L., Elce J.S., Jia Z.;
RT "Calpain mutants with increased Ca2+ sensitivity and implications for
RL the role of the C(2)-like domain.";
RL J. Biol. Chem. 276:7404-7407(2001).
RN [4]
RP MOTAGENESIS OF CYS-105; HIS-262; ASN-286 AND TRP-288.
RX MEDLINE-95361909; PubMed-7635186;
RA Arthur J.S., Gauthier S., Elce J.S.;
RT "Active site residues in m-calpain: Identification by site-directed
RL mutagenesis.";
RL FEBS Lett. 368:397-400(1995).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
RX MEDLINE-20069318; PubMed-10601010;
RA Hosfield C.M., Elce J.S., Davies P.L., Jia Z.;
RT "Crystal structure of calpain reveals the structural basis for
RL Ca(2+)-dependent protease activity and a novel mode of enzyme
RL activation.";
RL EMBO J. 18:6880-6889(1999).
CC -1- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which
CC catalyze limited proteolysis of substrates involved in
CC cytoskeletal remodelling and signal transduction.
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-I-Xaa, Met-I-Xaa or

```



```

CC -1- SUBUNIT: Heterodimer of a large (catalytic) and a small
CC (regulatory) subunit.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: CONTAINS 4 EF-HAND CALCIUM-BINDING DOMAINS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2; ALSO KNOWN AS THE
CC CALPAIN FAMILY OF THIOL PROTEASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X92523; CAA63301.1; -
DR EMBL; AF127766; AAD28255.2; -
DR HSSP; P04632; IDKV.
DR MEROPS; C02.004; -
DR MGD; MGI:107437; Capn3.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR001300; Peptidase_C2.
DR InterPro: IPR000169; Thiolprol_act_site.
DR Pfam; PF01067; Calpain_III; 1.
DR Pfam; PF00036; ethand; 3.
DR Pfam; PF00648; Peptidase_C2; 1.
DR PRINTS; PR00704; CALPAIN.
DR SMART; SM00230; Cyspc; 1.
DR SMART; SM00054; Efh; 2.
DR PROSITE; PS00018; EF_HAND; 2.
DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; FALSE NEG.
DR PROSITE; PS00640; THIOL_PROTEASE_ASN; FALSE NEG.
DR Hydrolase; Thiol protease; Calcium-binding; multigene family; Repeat;
DR Alternative splicing.
FT DOMAIN 49 234
FT DOMAIN 235 428 THIOL PROTEASE DOMAIN I.
FT DOMAIN 429 586 THIOL PROTEASE DOMAIN II.
FT DOMAIN 587 649 DOMAIN III, C2-LIKE DOMAIN.
FT DOMAIN 650 820 LINKER.
FT DOMAIN 821 942 EF-HAND 1 (PROBABLE).
FT CA_BIND 705 716 EF-HAND 2 (PROBABLE).
FT CA_BIND 735 746 ANCESTRAL CALCIUM SITE 3 (POTENTIAL).
FT DOMAIN 770 811 ANCESTRAL CALCIUM SITE 4 (POTENTIAL).
FT ACT_SITE 129 129 BY SIMILARITY.
FT ACT_SITE 334 334 BY SIMILARITY.
FT ACT_SITE 358 358 BY SIMILARITY.
FT VARSPLIT 268 315 MISSING (IN SHORT ISOFORM).
FT VARSPLIT 595 638 MISSING (IN SHORT ISOFORM).
SQ SEQUENCE 821 AA; 94213 MW; FF68A4E2C4355744 CRC64;

```

```

Query Match 16.68; Score 605.5; DB 1; Length 821;
Best Local Similarity 29.98; Pred. No. 2.5e-40;
Matches 167; Conservative 83; Mismatches 174; Indels 135; Gaps 18;

```

13 LFRDAAPPAASSTFCDSSTPLAOFREDITWRROEICATRLFPDDRECOVQKGLGCD 72
 14 LYLDEPPEPDSLTFYSQKFP1-QF---VKKRPEIENRPFITIGANRTDTCGGDGD 128
 73 CWFICACALLOKSRHLLDQVTPGQPSWADQYKGFCTRIQMGSGRWVEYTTDRLPCTA 132
 129 CWFIAAACLTLNRLERLFRVIPHDS--FTENVAGIFHFQWRMRGDMVDVYIDCLPYYN 186
 133 GRLEFSRQGRDEVFMLLEKVVAKVGSYHLMAGVADALVDLTGGLAERMLK----- 188
 187 NQVFTFSNHRNREWSALLERAKVAKLGSYALKGNTTEAMEDFTGVTFFEFKIDAPS 246
 189 -----GVAGSG----- 194
 247 DMYKMKRATERGSLMGCSIDDDGNNMTYGTGSPSLNMGELIARVNRNNDNSLRLDSLDIP 306

195 -GOODRPGWEHRTROLLHKDQCLISCVLSRPAAGRELGEPHATVSDRLQAG 253
 307 RGSDDR-----RTVPQYETRMAC-----GLVKHAASVGLLEALRK-G 348
 254 QCLILRIQNPWGRRCQGLMRGEGSGQVDAVAASELSQL-QGEGFWEEEFREF 312
 349 EKVLVALRNPWGQVENSGMSQDKMSFYDDEKARLQHOVTEDEFWSTIDFVYHF 408
 313 DELTVGPTVPEAGHLSLYERLCHTRAL-PCAMVKGASGGRNNGFP---SNPKF 367
 409 TKLEICNLTA-----LESDKLOTWTVSVNEGFWGSCAGGCRN---PPDFWTPQY 460
 368 MLNV-----SEPSEV---YIAVLQRSRRLHADMDAGRAALVGDSTHTSPSPISPKHY 417
 461 RLKLEEDDDPEDESVIYCSTFLVALMOKNR-----RKDRKL-----GANL 499
 418 QAVGLHLMKY-----EKRRVNLPRVLSMPVAGTACHAVDREYHLCESLPGYTLAVPS 471
 500 FTIGFAIYEPKREMGKQKLODPFLYNASKARSKTYINRREVSOFRPLPSEYTVPS 559
 472 TELKDAPEGFLRVFSTGR 490
 560 TYEPHOGEFILLRVFSEKR 578

RESULT 11
 CAN2_HUMAN
 ID CAN2_HUMAN STANDARD; PRT; 700 AA.
 AC P17655; Q16738;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Calpain 2, large (catalytic) subunit precursor (EC 3.4.22.17)
 DE (Calcium-activated neutral proteinase) (CANP) (M-type) (M-calpain)
 DE (Molluscal-calpain).
 GN CAPN2 OR CAPNP2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=89166474; PubMed=2852952;
 RA Imajoh S., Aoki K., Ohno S., Emori Y., Kawasaki H., Sugihara H.,
 RA Suzuki K.;
 RA "Molecular cloning of the cDNA for the large subunit of the
 RA high-Ca2+-requiring form of human Ca2+-activated neutral protease.";
 RA Biochemistry 27:8122-8128(1988).
 RN [2]
 RP SEQUENCE OF 1-79 FROM N.A.
 RC TISSUE=Lymph node;
 RX MEDLINE=89197947; PubMed=2539381;
 RA Hata A., Ohno S., Akita Y., Suzuki K.;
 RA "Randomly reiterated negative enhancer-like elements regulate
 RA transcription of a human gene for the large subunit of calcium-
 RA dependent protease.";
 RA J. Biol. Chem. 264:6404-6411(1989).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).
 RX MEDLINE=20105516; PubMed=10639123;
 RA Strobl S., Fernandez-Catalan C., Braun M., Huber R., Masumoto H.,
 RA Nakagawa K., Irie A., Sorimachi H., Bourenkov G., Bartunik H.,
 RA Suzuki K., Bode W.;
 RA "The crystal structure of calcium-free human m-calpain suggests an
 RA electrostatic switch mechanism for activation by calcium.";
 RA Proc. Natl. Acad. Sci. U.S.A. 97:588-592(2000).
 CC -1- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which
 CC catalyze limited proteolysis of substrates involved in
 CC cytoskeletal remodeling and signal transduction.
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-I-Xaa, Met-I-Xaa or
 CC Arg-I-Xaa with Leu or Val as the p2 residue.
 CC -1- COFACTOR: Binds 3 calcium ions.

- !- ENZYME REGULATION: Activated by 200-1000 micromolar concentrations of calcium and inhibited by calpastatin.
- !- SUBUNIT: Heterodimer of a large (catalytic) and a small (regulatory) subunit.
- !- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma membrane upon Ca++ binding.
- !- SIMILARITY: CONTAINS 5 EF-HAND CALCIUM-BINDING DOMAINS.
- !- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2; ALSO KNOWN AS THE CALPAIN FAMILY OF THIOL PROTEASES.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).

DR	EMBL; M23254;	AAA35645.1;	-;
DR	EMBL; J04700;	AAA52760.1;	-;
DR	PIR; A31218;	A31218.	
DR	PDB; IKFU; 07-DEC-01.		
DR	MEROPS; C02.002;	-;	
DR	MM; I14230;	-;	
DR	InterPro; IPR002048;	EF-hand.	
DR	InterPro; IPR001300;	Peptidase_C2.	
DR	InterPro; IPR000169;	Thiolprot_act_site.	
DR	Pfam; PF01067;	Calpain_III; 1.	
DR	Pfam; PF00036;	efhand; 3.	
DR	Pfam; PF00648;	Peptidase_C2; 1.	
DR	PRINTS; PR00704;	CALPAIN.	
DR	SMART; SMO0230;	Cyspc; 1.	
DR	SMART; SMO0054;	Efh; 2.	
DR	PROSITE; PS00018;	EF-HAND; 2.	
DR	PROSITE; PS00139;	THIOL_PROTEASE_CYS; 1.	
DR	PROSITE; PS00639;	THIOL_PROTEASE_HIS; FALSE.NEG.	
DR	PROSITE; PS00640;	THIOL_PROTEASE_ASN; FALSE.NEG.	
KW	Hydrolase; Thiol protease;	Calcium-binding; Repeat; Multigene family;	
KW	3D-structure.		
FT	PROPEP	1	19 ANCHORS TO THE SMALL SUBUNIT (POTENTIAL).
FT	CHAIN	20	700 CALPAIN 2, LARGE [CATALYTIC] SUBUNIT.
FT	DOMAIN	20	210 THIOL PROTEASE DOMAIN I.
FT	DOMAIN	211	355 THIOL PROTEASE DOMAIN II.
FT	DOMAIN	356	514 DOMAIN III, C2-LIKE DOMAIN.
FT	DOMAIN	515	529 LINKER.
FT	DOMAIN	530	700 DOMAIN IV.
FT	CA_BIND	541	552 EF-HAND 1.
FT	CA_BIND	585	596 EF-HAND 2.
FT	CA_BIND	615	626 EF-HAND 3.
FT	DOMAIN	650	661 ANCESTRAL CALCIUM SITE 3 (POTENTIAL).
FT	DOMAIN	680	691 ANCESTRAL CALCIUM SITE 4 (POTENTIAL).
FT	ACT_SITE	105	105 BY SIMILARITY.
FT	ACT_SITE	262	262 BY SIMILARITY.
FT	ACT_SITE	286	286 BY SIMILARITY.
FT	CONFLICT	68	68 S -> G (IN REF. 2).
FT	CONFLICT	73	74 MR -> IE (IN REF. 2).
SQ	SEQUENCE	700 AA;	80006 MW; A944D13BC8465531 CRC64;

[illegible]

Db	159	PTKDGELFLVHSAEGSEFWSALLEKAYAKINGCYEALSGATTEGPEFTGGIAEWYELK	218
QY	189	GVAGSGQODRPGRWEHRTCRQILLHLKDQCLISCCVLSPRAGARELGEF	242
Db	219	-----KPPNLFKIIQKA--LQKSLGSCSIDTSAADSEAITFQKLVKGHAYSV	266
QY	243	SDLRELQOGAGOCILLIRIONPWGRRCWGLNRGEGEGSOVDAAVASSELLSLOLEGFEW	302
Db	267	TGAEEVESN--GSLOKLIRINPWGEVETGRWVNDPCSWNTIDPEERELTRRHEGEFW	325
QY	303	VEEEFLUREFDELTVGPVTEAGHLSLYTERLLCHT-----RALPGAWVRQSGAGGR	356
Db	326	MSPSDFLRHYSRL-----EICNLTPDTLSDTYKKWKLTKMDGNWRRGSTAGGR	375
QY	357	N-NSGPSPNPKFWLRVSEPE-----VYIAVLQRSRLHAADWAGRALVGDSDHT	405
Db	376	NYPTFWANPYLIKLEEEDEDEEGSGCTFLVGLIQKHR-----PRQKMGEDMHT	428
QY	406	-SWSPASIPGKHYYQAVGLHWK----VEKRRVNLPRVLSMPPVAGTACHAYBREVHLRCE	460
Db	429	IGFGLIYVPEELSGQNIHLSKNFFLTNRABERSDTFINL-----REVNLRFK	476
QY	461	LSFGYILAVDPSTFKDAPGEFLLRVSTGRVLSAI	496
Db	477	LPPEGILVSTPEPNKDGDFCIRVSEKKADYQAV	512

RESULT 12

CA3	RAT	STANDARD;	PRT;	821	AA.
ID	CA3	RAT			
AC	PL6259;				
DT	01-AUG-1990	(Rel. 15, Created)			
DT	01-AUG-1990	(Rel. 15, Last sequence update)			
DT	16-OCT-2001	(Rel. 40, Last annotation update)			
DE	Calpain 3	large subunit (EC 3.4.22.17) (Calpain p94,			
DE	large [catalytic] subunit) (Calcium-activated neutral proteinase 3)				
DE	(CANP 3) (Muscle-specific calcium-activated neutral protease 3 large				
DE	subunit).				
GN	CANP3.				
OS	Rattus norvegicus (Rat).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
OX	NCBT_Taxid:10116;				
RN	[1]				
RC	SEQUENCE FROM N.A.				
RC	TISSUE=Skeletal muscle;				
RC	MEDLINE=90062125; PubMed=2555341;				
RA	Sorinachi H., Imajoh-Ohmi S., Emori Y., Kawasaki H., Ohno S.,				
RA	Minami Y., Suzuki K.;				
RT	"Molecular cloning of a novel mammalian calcium-dependent protease				
RT	distinct from both m- and mu-types. Specific expression of the mRNA				
RT	in skeletal muscle."				
RL	J. Biol. Chem. 264:20106-20111(1989).				
CC	!- FUNCTION: Calcium-regulated non-lysosomal thiol-protease.				
CC	!- CATALYTIC ACTIVITY: Preferential cleavage: Thr-l-Xaa, Met-l-Xaa or				
CC	Arg-l-Xaa with Leu or Val as the P2 residue.				
CC	!- ENZYME REGULATION: Activated by micromolar concentrations of				
CC	calcium and inhibited by calpastatin.				
CC	!- SUBUNIT: Heterodimer of a large (catalytic) and a small				
CC	(regulatory) subunit.				
CC	!- SUBCELLULAR LOCATION: Cytoplasmic.				
CC	!- TISSUE SPECIFICITY: SKELETAL MUSCLE.				
CC	!- MISCELLANEOUS: IN RAT THERE SEEMS TO BE 2 TYPES OF CALPAIN:				
CC	UBIQUITOUS FORMS - CALPAIN I (MICROMOLARS CA++ REQUIRING) AND				
CC	CALPAIN II (MILLI-MOLE CA++ REQUIRING), AND TISSUE SPECIFIC FORMS -				
CC	CALPAIN P94 AND NCL2. THE SMALL UNIT IS COMMON TO ALL FORMS.				
CC	!- SIMILARITY: CONTAINS 4 EF-HAND CALCIUM-BINDING DOMAINS.				
CC	!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2; ALSO KNOWN AS THE				
CC	CALPAIN FAMILY OF THIOL PROTEASES.				

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

Query Match	Similarity	16.4%	Score 599.5	DB 1	Length 821
Best Local	Similarity	31.0%	Pred. No. 7,4e-40		
Matches	170	Conservative	78	Mismatches 186	Indels 115
					Gaps 18
QY	13	LFRDAFPAADSSLFCDUSTPLAQRREDITWRPOECATPRLFPDDPREGVKOGSLGD	72		
Db	74	LYLDEPFPPDESLTSYQKFPi-QE----	VWKRPEPICENPFFIIGGARTRDICOGDLPLD	128	
QY	73	CMFLCACAALOKSRHLDOVITPPOGPSNADQYRGSFICRIOMQFGRWVYTTDDDLPCLA	132		
Db	129	CMFLAALACLTINELLRLRVRIPHDOS--FTEYAGIHFHFQFRRYGDWDVYIDDCIPTYN	186		
QY	133	GRLCESRCQREDVEWMLPFLLEKYAVFVHGSYEHLAMAGOVADALVDYLTGGLAERMNK	188		
Db	187	NOLVFTKSNHNRNEEFASALLEKAYAKLHSGSYEALKKGNTTEAMEDPTGGVTEFEELKADPS	246		
QY	189	-----GVAGSGGQODRPGRWERTCQQLHLKDCCLT	220		
Db	247	DMYTKMRALIERGSLMGCSIDDTGNTMYTGSTSSGLNM--GELIARWYNNM----	DNSLT	299	
QY	221	SCCVLSPRAG-----ARELGEFFAFTVSDRELQAGOGACITLLRIQN	263		
Db	300	RDSDDLPDRAASDDPSRTIVPQVETFRMACGLVKGHAIVYTGLEALFK-GEKVKLVLRLN	358		
QY	264	PMGRRCOGGLRREGGEGSSOVDAANAASLLLSQL-QEGEFTWVEEELFLRFDELVTGYPT	322		
Db	359	PMGOVFNWNGSWDKDMSFVDEKDEKARLQHOVEDGEGFMSYDDFVYHFTLIECLNTLA	418		
QY	333	EAGHLQSLYTRRLCLHTRAL-PGAMVKKQSAGGCRNNNSGFP-----SNPREFMLRYSE---	373		
Db	419	DA-----LESQKIQWYTVSVNEGRRVRCGSAAGCCN--FPDTFTNTQYRIKLKLEDDDD	470		
QY	374	--PSEY-----YIAYLQSRRLAADMGARALVGDHSITWSBPASTPGKHAYQVGLHMKV	427		
Db	471	PDDSEVCSFTVALAKQKR-----RDRKL-----GANLFTIGTAIYEV	509		

	RESULT	13
ID	CAN2_CHICK	
AC	CAN2_CHICK	STANDARD: PRT: 700 AA.
DT	092178;	
DT	01-NOV-1997	(Rel. 35, Created)
DT	01-NOV-1997	(Rel. 35, Last sequence update)
DT	16-OCT-2001	(Rel. 40, Last annotation update)
DE	Calpain 2, large [catalytic] subunit precursor (EC 3.4.22.17)	
DE	(Calcium-activated neutral proteinase) (CANP) (M-type) (M-calpain)	
DE	(Millimolar calpain).	
GN	CAPN2.	
OS	Gallus gallus (Chicken).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;	
CC	Gallus.	
CC	NBI_TaxID=9031;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RP	TISSUE=MUSCLE;	
RC	MEDLINE=95260862; PubMed=7742367;	
RX	Sornachi H., Tsukahara T., Okada-Ban M., Sugita H., Ishiura S.,	
RA	Suzuki K.;	
RT	"Identification of a third ubiquitous calpain species -- chicken	
RT	musclic expresses four distinct calpains.";	
RL	Biochim. Biophys. Acta 1261:381-393(1995).	
CC	-1- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which	
CC	catalyze limited proteolysis of substrates involved in	
CC	cytoskeletal remodeling and signal transduction (by similarity).	
CC	-1- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-I-Xaa, Met-I-Xaa or	
CC	Arg-I-Xaa with Leu or Val as the P2 residue.	
CC	-1- COFACTOR: Binds 3 calcium ions.	
CC	-1- ENZYME REGULATION: Activated by 200-1000 micromolar concentrations	
CC	of calcium and inhibited by calpastatin.	
CC	-1- SUBUNIT: Heterodimer of a large (catalytic) and a small	
CC	(regulatory) subunit.	
CC	-1- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma	
CC	membrane upon Ca++ binding.	
CC	-1- SIMILARITY: CONTAINS 5 EF-HAND CALCIUM-BINDING DOMAINS.	
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2; ALSO KNOWN AS THE	
CC	CALPAIN FAMILY OF THIOL PROTEASIS.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC	or send an email to license@isb-sib.ch).	

DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR001300; Peptidase_C2.
DR InterPro: IPR000169; Thiolprot_act_site.
DR Pfam: PF01067; Calpain_III; 1.
DR Pfam: PF00036; efband_3.
DR Pfam: PF00648; Peptidase_C2; 1.
DR PRINTS; PRO0704; CALPAIN.
DR SMART; SM00230; Cyspec; 1.
DR SMART; SM00054; Efn; 2.
DR PROSITE; PS00018; EF_HAND; 2.
DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.

DR	PROSITE; PS00639; THIOL-PROTEASE HIS; FALSE-NEG.	
DR	PROSITE; PS00640; THIOL-PROTEASE-ASN; FALSE-NEG.	
KW	Hydrolase; Thiol protease; Calcium-binding; Multigene family.	
FT	PROPEP 1 19 ANCHORS TO THE SMALL SUBUNIT (POTENTIAL).	
FT	CHAIN 20 700 CALPAIN 2, LARGE [CATALYTIC] SUBUNIT.	
FT	DOMAIN 20 210 THIOL-PROTEASE DOMAIN I.	
FT	DOMAIN 211 355 THIOL-PROTEASE DOMAIN II.	
FT	DOMAIN 356 514 THIOL-PROTEASE DOMAIN I.	
FT	DOMAIN 515 529 LINKER.	
FT	DOMAIN 530 700 DOMAIN IV.	
FT	CA_BIND 541 552 EF-HAND 1.	
FT	CA_BIND 585 596 EF-HAND 2.	
FT	CA_BIND 615 626 EF-HAND 3.	
FT	DOMAIN 650 661 ANCESTRAL CALCIUM SITE 4 (POTENTIAL).	
FT	DOMAIN 680 691 ANCESTRAL CALCIUM SITE 5 (POTENTIAL).	
FT	ACT_SITE 105 105 BY SIMILARITY.	
FT	ACT_SITE 262 262 BY SIMILARITY.	
FT	ACT_SITE 286 286 BY SIMILARITY.	
SQ	SEQUENCE 700 AA; 79228 MW; C3AEDB39CCB56D3B CRC64;	
Query Match		
Best Local Similarity 16.4%; Score 597.5; DB 1; Length 700;		
Matches 163; Conservative 81; Mismatches 203; Indels 75; Gaps 17;		
QY	1 MRAGRGATPARELFRDAAPDAADSLFCDLSTPLAQFREDITWRRPQEICATPRLPFPDDP 60	
DB	40 LOGGR-----LFHDPSPAGPAALGYRELGNPKYKGVVCRPTELSCPRFTAGCA 92	
QY	61 REGOVKGLGDCWFLCAALOKSHLDVIPPQPSWADOEYKSTCRIMQFGRWV 120	
DB	93 TRTDICOGALGDCWLLAAASLTLEILARVVRDQ-SFQD-EVAGIPHFQWQYGEWV 150	
QY	121 EVTTDRPLCLAGRCFSRCQREDVFWLPLEKRVYAKVGSVEHLWAGOVADALYDLTGG 180	
DB	151 DVVVDRPLTKGELLFVISAEGSEFWALKAYAKLNGSYEALSGGTTTGGFEDFTGG 210	
QY	181 LAERNWLVAGSGGQDPRGWEHRTCRQLLH--LKDQCLISCCVLSPRAG-----AR 232	
DB	211 IAEWYEL-----QKAPNLF-----KIQKALQKSLGCSIDITSAAETAVTSQ 256	
QY	233 ELGEFHAFIVSDLRLOGAGQCIILLRTONPWRCWQCLRWEGEGESQVDAVASSEL 292	
DB	257 KLKVGHAYSVTGAEEVNF-RGSIQKLRIRNPGEVETGKWNDCNPNMSGVDPEVRRL 315	
QY	293 LSLQGEFEWVEEELREFDELTVG--YPVTEAGHLQSLYTERLLCHTRALPGAWVQ 350	
DB	316 TRRHEDGEFWMAFNDELRYHSRLCINTLPTDLASDRYKNS--LL-----KLDGNWRRGA 369	
QY	351 SAGGCNR-NSGSPSPKFWLVSEPS-----VYIAYLQSLRHLAADWAGRARAL 399	
DB	370 TAGGCNYPNTFTWNPQYLKLEEDDEDDPDEGGCTFLGLIQKHR-----RKQREM 422	
QY	400 VGDST-SWSPASIPGKHQVAVGLHLWK-----VEKRVNLPVLVSPVPPVAGTACHAYDRE 454	
DB	423 GEDMHTGFIAYEVPFESGQNIHLKSNFFLTNAREKSNFTINL-----RE 470	
QY	455 VHLRCLSPGYLVAPSTPLKDPAGEFLRVFSTGRVLSAI 496	
DB	471 VLNRRFLPAGEIIVPSTPEPNLNGDFCLRVSEKNANSTVI 512	
RESULT 14		
CAN3_CHICK		
ID	CAN3_CHICK STANDARD; PRT; 810 AA.	
AC	Q92177;	
DT	01-NOV-1997 (Rel. 35, Created)	
DT	01-NOV-1997 (Rel. 35, Last sequence update)	
DT	16-OCT-2001 (Rel. 40, Last annotation update)	
DE	Calpain 3 large subunit (EC 3.4.22.17) (Calpain L3) (Calpain p94,	
DE	large [catalytic] subunit) (Calcium-activated neutral proteinase 3)	
DE	(CANP 3) (Muscle-specific calcium-activated neutral protease 3 large	
DE	subunit).	

GN	CAPN3.	
OS	Gallus gallus (Chicken).	
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianinae;	
OC	Gallus.	
OX	NCBI_TaxID=9031;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=Lung;	
RX	MEDLINE=95260862; PubMed=7742367;	
RA	Sorimachi H., Tsukahara T., Okada-Ban M., Sugita H., Ishiura S.,	
RA	Suzuki K.;	
RT	*Identification of a third ubiquitous calpain species -- chicken	
RT	muscle expresses four distinct calpains.*;	
RL	Biochim. Biophys. Acta 1261:381-393(1995).	
CC	-!- FUNCTION: Calcium-regulated non-lysosomal thiol-protease.	
CC	-!- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-I-Xaa, Met-I-Xaa or	
CC	Arg-I-Xaa with Leu or Val as the P2 residue.	
CC	-!- ENZYME REGULATION: Activated by micromolar concentrations of	
CC	calcium and inhibited by calpastatin.	
CC	-!- SUBUNIT: Heterodimer of a large (catalytic) and a small	
CC	(regulatory) subunit.	
CC	-!- SUBCELLULAR LOCATION: Cytoplasmic and nuclear (Potential).	
CC	-!- TISSUE SPECIFICITY: SKELETAL MUSCLE. LOW LEVELS IN SPLEEN,	
CC	INTESTINE AND BONE.	
CC	-!- SIMILARITY: CONTAINS 4 EF-HAND CALCIUM-BINDING DOMAINS.	
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2; ALSO KNOWN AS THE	
CC	CALPAIN FAMILY OF THIOL PROTEASES.	

This SWISS-PROT entry is copyright. It is produced through a collaboration		
between the Swiss Institute of Bioinformatics and the EMBL outstation -		
the European Bioinformatics Institute. There are no restrictions on its		
use by non-profit institutions as long as its content is in no way		
modified and this statement is not removed. Usage by and for commercial		
entities requires a license agreement (See http://www.isb-sib.ch/announce/		
or send an email to license@sib-sib.ch).		

DR	EMBL; D38028; BAA07230.1; -	
DR	HSSP; P04574; ITALY.	
DR	MEROPS; C02.004; -	
DR	InterPro; IPR002048; EF-hand.	
DR	InterPro; IPR001300; Peptidase C2.	
DR	InterPro; IPR000169; Thiolprot_act_site.	
DR	Pfam; PF01067; Calpain.III; 1.	
DR	Pfam; PF00036; ehand; 3. 2.	
DR	Pfam; PF00648; Peptidase_C2; 1.	
DR	PRINTS; PR00704; CALPAIN.	
DR	SMART; SM00230; Cyspc; 1.	
DR	SMART; SM00054; EFh; 2.	
DR	PROSITE; PS0018; EF HAND; 2.	
DR	PROSITE; PS00139; THIOL-PROTEASE_CYS; 1.	
DR	PROSITE; PS00639; THIOL-PROTEASE_HIS; FALSE-NEG.	
DR	PROSITE; PS00640; THIOL-PROTEASE-ASN; FALSE-NEG.	
KW	Hydrolase; Thiol protease; Calcium-binding; Nuclear protein;	
KW	Multigene family.	
FT	DOMAIN 43 228 THIOL-PROTEASE DOMAIN I.	
FT	DOMAIN 229 421 THIOL-PROTEASE DOMAIN II.	
FT	DOMAIN 422 579 DOMAIN III, C2-LIKE DOMAIN.	
FT	DOMAIN 580 638 LINKER.	
FT	DOMAIN 639 809 DOMAIN IV.	
FT	CA_BIND 694 705 EF-HAND 1 (PROBABLE).	
FT	CA_BIND 724 735 EF-HAND 2 (PROBABLE).	
FT	DOMAIN 759 770 ANCESTRAL CALCIUM SITE 3 (POTENTIAL).	
FT	DOMAIN 789 800 ANCESTRAL CALCIUM SITE 4 (POTENTIAL).	
FT	ACT_SITE 123 123 BY SIMILARITY.	
FT	ACT_SITE 327 327 BY SIMILARITY.	
FT	ACT_SITE 351 351 BY SIMILARITY.	
SQ	SEQUENCE 810 AA; 93560 MW; E8DE99411C51041E CRC64;	

Query Match 16.2%; Score 589.5; DB 1; Length 810;
Best Local Similarity 28.6%; Pred. No. 4.5e-39;
Matches 161; Conservative 87; Mismatches 178; Indels 137; Gaps 19;

RL Mol. Cell. Biol. 15:824-834(1995).
 RP SEQUENCE FROM N.A. (LONG ISOFORM).
 RC STRAIN-BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blaise R.G., Champagne M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,
 RA Abvill J.F., Abayyan A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Bailev R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borrova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Chew S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durkin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Fouts C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeigwan C.,
 RA Jambal B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Matzel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Plittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Renkert K., Remington K., Saunders R.D.C., Scheibel F., Shen H.,
 RA Shue B.C., Siden-Kimms I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E.C., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.M., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 CC -1- FUNCTION: CALPAINS ARE CALCIUM-ACTIVATED NON-LYSOSOMAL THIO-
 CC PROTEASES. INVOLVED IN THE ORGANIZATION OF THE ACTIN-RELATED
 CC CYTOSKELETON DURING EMBRYOGENESIS.
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-[Xaa, Met-[Xaa or
 CC Arg-[Xaa with Leu or Val as the P2 residue.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS. LONG (SHOWN HERE) AND
 CC SHORT. ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED. HIGHEST IN THE CENTRAL
 CC NERVOUS SYSTEM, BLOOD CELLS AND MIDGUT.
 CC -1- MISCELLANEOUS: THIS PROTEIN BINDS CALCIUM.
 CC -1- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2; ALSO KNOWN AS THE
 CC CALPAIN FAMILY OF THIOL PROTEASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X78555; CAAS5298.1; -
 DR EMBL: X78555; CAAS5297.1; ALT_INIT.
 DR EMBL: Z46891; CAAB6993.1; -
 DR EMBL: Z46892; CAAB6994.1; -
 DR EMBL: AE003796; AAF57563.1; -
 DR HSSP: P17655; IDKV.
 DR MEROPS: C02.014; -.

DR FlyBase: FBgn0012051; Calpa.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR001300; Peptidase_C2.
DR InterPro: IPR001067; Thiolprot_act_site.
DR Pfam: PF01067; Calpain III; 1.
DR Pfam: PF00036; efhand; 2.
DR Pfam: PF00648; Peptidase_C2; 1.
DR PRINTS: PR00704; CALPAIN.
DR SMART: SM00230; Cyspc; 1.
DR SMART: SM00054; EFh; 2.
DR PROSITE: PS00018; EF_HAND; 1.
DR PROSITE: PS00139; THIOL_PROTEASE_CYS; 1.
DR PROSITE: PS00639; THIOL_PROTEASE_HIS; FALSE_NEG.
DR PROSITE: PS00640; THIOL_PROTEASE_ASN; FALSE_NEG.
KW Hydrolase; Thiol protease; Calcium-binding; Multigene family;
KW Alternative splicing.
FT DOMAIN 68 248 THIOL_PROTEASE DOMAIN I.
FT DOMAIN 249 402 THIOL_PROTEASE DOMAIN II.
FT DOMAIN 403 557 DOMAIN III, C2-LIKE DOMAIN.
FT DOMAIN 558 577 LINKER.
FT DOMAIN 578 827 DOMAIN IV.
FT CA_BIND 712 723 EF_HAND 1 (PROBABLE).
FT CA_BIND 742 753 EF_HAND 2 (PROBABLE).
FT DOMAIN 777 789 ANCESTRAL CALCIUM SITE 3 (POTENTIAL).
FT DOMAIN 807 819 ANCESTRAL CALCIUM SITE 4 (POTENTIAL).
FT ACT_SITE 143 143 BY SIMILARITY.
FT ACT_SITE 299 299 BY SIMILARITY.
FT ACT_SITE 327 327 BY SIMILARITY.
FT VARSPLIC 554 558 ENDDH -> RTSRQ (IN SHORT ISOFORM).
FT VARSPLIC 559 828 MISSING (IN SHORT ISOFORM).
FT CONFLICT 35 35 I -> Y (IN REF. 2).
FT CONFLICT 306 306 V -> I (IN REF. 2).
FT CONFLICT 397 397 N -> H (IN REF. 1).
SQ SEQUENCE 828 AA; 93962 MW; 09576D1268BD569C CRC64;

Search completed: September 19, 2002, 14:54:24
Job time: 263 sec

Query Match 16.1%; Score 586; DB 1; Length 828;
Best Local Similarity 30.3%; Pred. No. 8.8e-39;
Matches 167; Conservative 83; Mismatches 185; Indels 116; Gaps 22;
QY 13 LFRDAAPPAADSSIFCDLSTPLAQF--RED--ITWRPQIECATPLRPDDPREGQVKG 68
DB 88 LFEDPLFPASNESL-----QFSRPRDHIEWLPHEIAENPQFFVEGYSRFDVQQG 138
QY 69 LIGDCHFLCACAALQKSRHLLDQVIPPQPSWADQYRGSTCRINQFGRWVEVTTDDR 128
DB 139 ELGDCWLLAATANLTQESNLFRRVPAEQS--FEENYAGIFHFYQYKWDVVIDDL 196
QY 129 PCLAGRLCFKSCQREDVFLPLEKVTAKYVGSYHLWAGOVADALVDLTGGLAERNLK 188
DB 197 PTYNGELMYHSTERNKFEWSALLEKAYAKLHGSYALKGGSTCEAMEDFTGGVSEWYDLK 256
QY 189 GVAGS-----GGQDRPGRWEHRTCRQLHLKDKQLISCCVLSPRAGAR 232
DB 257 EAPGNLFTILQKAAERNMGCSEPDNPVTEAETPQGLIRGHAYSITKVLID----- 310
QY 233 ELGEFHAFIVSDRLQCGAQCCILLRIONPWGRRC-WOGLWREGGEGWSQVDAVASE 291
DB 311 -----IVTPNR--QKG---IPMIRMRPNWNEAWNGPWSDDSPWRYIPEQKAE 356
QY 292 LLSQLOGEFWEFEERFLERDELTV---GYPTVE---AGHLQ---SLYTERLLCHT 339
DB 357 IGLTFDRDGEFMSFQDLNHFDRVEICNLSPDSLTDQONSGRKRWMSY----- 408
QY 340 RALPGAWKQSGAGGCRN-NSGFPSNPKFWLRVSEPE-----VYIAVLQSRSLHAA 390
DB 409 ---EGEWTPGVTAGGCRNFLTDFHNPQYIITLVDPDEDEEGQCIVYVALMOKNR---- 461
QY 391 DWAGRARALVGDHSTNSPASIPGKHQYQVGLHLMVKERRV-NLPRVLSM---PPVAGT 446
DB 462 -----RSKRN-----GMECLTIGFAIYSLNDRLENRPQGLNFRYKSSVGR 504
QY 447 ACHAYD-REVHLRCELSPGYVLAAPPSTFLKDAFGELLRVFSTGRVSL-----AIRAV 499

Thu Sep 19 15:10:44 2002

us-09-768-877-2.sepl9.rsp

OM protein - protein search, using sw model

Run on: September 19, 2002, 14:48:56 ; Search time 21.45 Seconds

(without alignments)
3010.351 Million cell updates/sec

Title: US-09-768-877-2

sequence: 1 MRAGRGATPARELFRDAFP.....SQEMLGQFLQEVSVMAVMKT 672

Scoring table: BLOSUM62

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum match 08

Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*

```
1:  plr1: *
2:  plr2: *
3:  plr3: *
4:  plr4: *
```

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	642	17.6	705	1	C1CHHH	calpain (EC 3.4.22)
2	618.5	16.9	714	1	C1HHH	calpain (EC 3.4.22)
3	618	16.9	821	1	C1HHH3	calpain (EC 3.4.22)
4	614	16.8	700	2	S38361	calpain (EC 3.4.22)
5	612	16.8	703	2	A48764	calpain (EC 3.4.22)
6	608	16.7	700	1	C1HHH2	calpain (EC 3.4.22)
7	599.5	16.4	821	1	B34488	calpain (EC 3.4.22)
8	597.5	16.4	700	2	S57194	calpain (EC 3.4.22)
9	589.5	16.2	810	1	S57196	calpain (EC 3.4.22)
10	588	16.1	805	1	A55054	calpain (EC 3.4.22)
11	579	15.9	653	1	A44749	calpain (EC 3.4.22)
12	569.5	15.6	647	2	UC5772	calpain (EC 3.4.22)
13	565	15.6	648	2	S71885	calpain (EC 3.4.22)
14	540	14.8	758	1	A39343	calpain (EC 3.4.22)
15	506.5	13.9	586	2	S57195	calpain (EC 3.4.22)
16	457.5	12.5	2143	2	G96595	calpain (EC 3.4.22)
17	376.5	10.3	1557	1	BVPSL	hypothetical protein
18	376.5	10.3	1557	2	T24825	hypothetical protein
19	335.5	9.12	616	2	T24825	hypothetical protein
20	326	8.9	634	2	T33871	hypothetical protein
21	271.5	7.4	805	2	T24431	hypothetical protein
22	268.5	7.4	422	2	B24815	hypothetical protein
23	191.5	5.2	1134	2	C88040	calpain (EC 3.4.22)
24	189.5	5.2	617	2	T21175	calpain (EC 3.4.22)
25	162.5	4.5	623	2	T21177	hypothetical protein
26	120.5	3.3	217	2	T26132	hypothetical protein
27	114	3.1	499	1	C44767	benzoylformate decarboxylase
28	111	3.0	757	2	C84120	subtilisin-type protease
29	110	3.0	3519	2	S43048	polyketide synthase

30	107	2.9	741	2	G84888	probable transketolase
31	107	2.9	1200	2	T35559	probable DNA methyltransferase
32	106	2.9	761	2	UC5759	brain-specific serine/threonine protein kinase
33	105.5	2.9	687	2	S19680	protein-glutamine gamma-glutamyl transferase
34	104.5	2.9	673	2	T00328	hypothetical protein
35	102.5	2.8	333	2	S55593	connexin 37 - human
36	102.5	2.8	667	1	A39045	protein: glutamine gamma-glutamyl transferase
37	102.5	2.8	744	2	A82822	MDM2-ubiquitin ligase
38	102	2.8	573	2	JC4335	anti-mullerian hormone
39	102	2.8	2504	1	A57788	fatty-acid synthase
40	101.5	2.8	330	2	A24815	calpain (EC 3.4.22.2)
41	101.5	2.8	611	2	A11941	hypothetical protein
42	101	2.8	501	2	AH1852	hypothetical protein
43	100.5	2.8	1453	1	A53210	phospholipase A2
44	99.5	2.7	644	2	UC5119	anti-mullerian hormone
45	99.5	2.7	1487	2	S48719	phospholipase A(2)

RESULT 1

CLC:HH

calpain (EC 3.4.22.17) large chain 4 - chicken

C;Species: Gallus gallus (chicken)

Ci:Accession: A00979
#sequence_revision 09-Aug-1997 #text_change 24-Nov-1999
C:Date: 1/-May-1985

R. Ohno, S.; Emori, Y.; Imajob, S.; Kawasaki, H.; Katsaragi, M.; Suzuki, K. *Nature* 312, 566-570, 1984

A; Title: Evolutionary origin of a calcium-dependent protease by fusion of genes for
A; Reference number: A93348; MUID:85061606
A; Accession: 000070

A: Molecule type: mRNA

A;Residues: 1-705 <OHN
A;Cross-references: EM

R. Remori, Y.; Ohno, S.; Tobita, M.; Suzuki, K.
FEBS Lett. 194, 249-252, 1986

A:Reference number: A91354; MUID:86082358
A:Contents: annotation, gene structure

C;Comment: This calpain has calcium requirements intermediate between those of the

C;Genetics: 82/3; 106/1; 145/3; 190/2; 246/3; 274/3; 303/2; 328/2; 382/1; 438/3; 442

C:Complex: heterodimer of L (large) and S (small) chains
C:Function:

A;Description: catalyzes the hydrolysis of peptides

C/Superfamily: calpain large chain; calmodulin repeat homology; calpain catalytic domain; note: cleaves preferentially after tyrosine, methionine, or arginine residues and

Keywords: blocked amino end; calcium binding; cysteine proteinase; duplication; E

F;533-564/Domain: calmodulin repeat homology <EF1>
E:576-608/Domain: calmodulin repeat homology <EF1>

F; 609-638/Domain: calmodulin repeat homology <EF3>
F; 609-638/Domain: calmodulin repeat homology <EF2>
F; 609-638/Domain: calmodulin repeat homology <EF1>
F; 609-638/Domain: calmodulin repeat homology <EF0>

F:674-705/Domain: calmodulin repeat homology <EF4>
E:641-673/Domain: calmodulin repeat homology <EF5>

F;2/Modified site: blocked amino end (Met) (in mature form) #status experimental
F;108,265,289/Active site: Cys, His, Asn #status predicted

Country	Year	Population (millions)	Urban population (millions)	Urban population (%)	Population density (per sq km)	Urban population density (per sq km)	Population growth rate (%)	Urban population growth rate (%)	Population growth rate (%)	Urban population growth rate (%)	Population growth rate (%)	Urban population growth rate (%)
Algeria	1980	10.0	4.0	40.0	100	400	1.5	2.5	1.5	2.5	1.5	2.5
Algeria	1985	10.5	4.5	42.9	105	450	1.8	2.8	1.8	2.8	1.8	2.8
Algeria	1990	11.0	5.0	45.5	110	500	2.0	3.0	2.0	3.0	2.0	3.0
Algeria	1995	11.5	5.5	47.8	115	550	2.2	3.2	2.2	3.2	2.2	3.2
Algeria	2000	12.0	6.0	50.0	120	600	2.5	3.5	2.5	3.5	2.5	3.5
Algeria	2005	12.5	6.5	52.0	125	650	2.8	3.8	2.8	3.8	2.8	3.8
Algeria	2010	13.0	7.0	53.8	130	700	3.0	4.0	3.0	4.0	3.0	4.0
Algeria	2015	13.5	7.5	55.6	135	750	3.2	4.2	3.2	4.2	3.2	4.2
Algeria	2020	14.0	8.0	57.1	140	800	3.5	4.5	3.5	4.5	3.5	4.5
Algeria	2025	14.5	8.5	58.6	145	850	3.8	4.8	3.8	4.8	3.8	4.8
Algeria	2030	15.0	9.0	60.0	150	900	4.0	5.0	4.0	5.0	4.0	5.0
Algeria	2035	15.5	9.5	61.3	155	950	4.2	5.2	4.2	5.2	4.2	5.2
Algeria	2040	16.0	10.0	62.5	160	1000	4.5	5.5	4.5	5.5	4.5	5.5
Algeria	2045	16.5	10.5	63.6	165	1050	4.8	5.8	4.8	5.8	4.8	5.8
Algeria	2050	17.0	11.0	64.7	170	1100	5.0	6.0	5.0	6.0	5.0	6.0
Algeria	2055	17.5	11.5	65.7	175	1150	5.2	6.2	5.2	6.2	5.2	6.2
Algeria	2060	18.0	12.0	66.7	180	1200	5.5	6.5	5.5	6.5	5.5	6.5
Algeria	2065	18.5	12.5	67.6	185	1250	5.8	6.8	5.8	6.8	5.8	6.8
Algeria	2070	19.0	13.0	68.4	190	1300	6.0	7.0	6.0	7.0	6.0	7.0
Algeria	2075	19.5	13.5	69.2	195	1350	6.2	7.2	6.2	7.2	6.2	7.2
Algeria	2080	20.0	14.0	70.0	200	1400	6.5	7.5	6.5	7.5	6.5	7.5
Algeria	2085	20.5	14.5	70.7	205	1450	6.8	7.8	6.8	7.8	6.8	7.8
Algeria	2090	21.0	15.0	71.4	210	1500	7.0	8.0	7.0	8.0	7.0	8.0
Algeria	2095	21.5	15.5	72.1	215	1550	7.2	8.2	7.2	8.2	7.2	8.2
Algeria	2100	22.0	16.0	72.7	220	1600	7.5	8.5	7.5	8.5	7.5	8.5
Algeria	2105	22.5	16.5	73.3	225	1650	7.8	8.8	7.8	8.8	7.8	8.8
Algeria	2110	23.0	17.0	73.9	230	1700	8.0	9.0	8.0	9.0	8.0	9.0
Algeria	2115	23.5	17.5	74.5	235	1750	8.2	9.2	8.2	9.2	8.2	9.2
Algeria	2120	24.0	18.0	75.0	240	1800	8.5	9.5	8.5	9.5	8.5	9.5
Algeria	2125	24.5	18.5	75.5	245	1850	8.8	9.8	8.8	9.8	8.8	9.8
Algeria	2130	25.0	19.0	76.0	250	1900	9.0	10.0	9.0			

Query Match	Score	DB 1;	Length
17.68;	642;	DB 1;	Length 705;

Best local similarity 32.0%; Pred. No. 2.1e-42;
Matches 162; Conservative 78; Mismatches 188; Indels 78; Gaps 16;

13 LERDAEPADSSL-FCDLSTPLAOFREDITWRBPOTCATPDELDDNBECOVKOCITC 71

[illegible]

10 MANDYK ET AL. 1995

```

0Y      / 2 DCWFLCACALQKSRHLDDVPIPGQPSWADQFYRGSFTRCWIQGRWEVTYDDRLPCL 131
||| || : | : || || :: || || || || || || || || || || || || || || || || ||

```

107 DCWLLAIGSLTNEELLHRYVPHGQS--FQEDYAGIFHFQIQGCEWVDVVDDLLPTK 164

132 AGRICFSCQREDVEFWLPLEKYAKVHGSEHLMAGQVADALVDLTGGLAERMNLKGVA 191

Db 165 DGEVLFVHSAECTFWFSALEKAYAKNGCYESLSGGSTTEGFEFTGGVAMVDLK --- 221
QY 192 GSGGQDQDRPGRWEHRTCRQLLH ----- LKQCLISCVLSPRAGARELGEF ----- HAF 240
Db 222 ----- RAPRNGHIIIRKALERSLLGCSIDITSADFMEAVTFKKLVKCHAY 267
QY 241 IVSDLREL--QGAGQCILLRTONPWRCRCWGLWRGEGSGVQVDAVASSELLSOLQE 298
Db 268 SVTAFKDVNTVNGQEQ---LIRINPWGQVEMTGKSGSDSEWDNIDPDSREELQLKMD 324
QY 299 GEFWVEEELREFDELTVGYPVTEAGHLQSLYTERL-LCHTRALPCAWYKQSGAGCGRN 357
Db 325 GEFWMSRDFWRFESRLEICNLTPDA-----LTKDELSRWHTQVFEGTWRGSGTAGCGRN 379
QY 358 N-SGFPNPFKWLVRVSEPE-----VYIAVLQSRSLHAADWAGRALVGDSTH- 405
Db 380 NPATFWINPQFKILLEDDDDGDEVACSFYVALMQKHR-----RRRRVGGDMHTI 432
QY 406 SWSPASTPGHYQAVGLHLWK-----VEKRVNLPVLSMPVAVGTACHAYDREVHLRCEL 461
Db 433 GFAYIEVPEAOGSQNVHLKDKDFLRNQSRARSETFINL-----REVSNQIRL 480
QY 462 SPGYLAVPSTELKDPAGEFLLRVES 487
Db 481 PGEYIIVPSTFEFHKHEADFILRVFT 506
RESULT 2
CIHUH3
N:Alternate names: calpain chain 1 [validated] - human
N:Contents: chemoattractant factor
C:Species: Homo sapiens (man)
C:Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 08-Dec-2000
C:Accession: A36740; S10591
R:Aoki, K.; Imajoh, S.; Ohno, S.; Emori, Y.; Koike, M.; Kosaki, G.; Suzuki, K.
FEBS Lett. 205, 313-317, 1986
A:Title: Complete amino acid sequence of the large subunit of the low-Ca2+-requiring for
A:Reference number: A26213; MUID:86301172
A:Accession: A26213
A:Molecule type: mRNA
A:Residues: 1-714 <OK>
R:Cross-references: EMBL:X04366; NID:g29663; PIDN:CAA27881.1; PID:g29664
R:Kunimatsu, M.; Higashiyama, S.; Sato, K.; Ohkubo, I.; Sasaki, M.
Biochem. Biophys. Res. Commun. 164, 875-882, 1989
A:Title: Calcium dependent cysteine proteinase is a precursor of a chemotactic factor for
A:Reference number: A36740; MUID:90056492
A:Accession: A36740
A:Molecule type: protein
A:Residues: 2-10 <KUN>
R:Experimental source: erythrocytes
R:Sorimachi, H.; Ohmi, S.; Emori, Y.; Kawasaki, H.; Saido, T.C.; Ohno, S.; Minami, Y.; S
Biol. Chem. Hoppe-Seyler 371(Suppl.), 171-176, 1990
A:Title: A novel member of the calcium-dependent cysteine protease family.
A:Reference number: S10589; MUID:90380278
A:Contents: annotation
A:Note: comparison with other gene products
C:Comment: Calpain I is activated by micromolar concentrations of calcium.
C:Genetics:
A:Gene: GDB:CAPN1; mu-CANP
A:Cross-references: GDB:119749; OMIM:114220
A:Map position: 11pter-11qter
A:Complex: heterodimer of L (large) and S (small) chains
C:Function:
A:Description: catalyzes the hydrolysis of peptides
A:Note: cleaves preferentially after tyrosine, methionine, or arginine residues and before
C:Superfamily: calpain large chain; calmodulin repeat homology; calpain catalytic domain
C:Keywords: acetylated amino end; calcium binding; cysteine proteinase; duplication; EF
F:2-714/Product: calpain large chain 1 #status predicted <NAT>
F:2-10/Product: chemotactic factor #status experimental <CHF>
F:85-337/Domain: calpain catalytic domain homology <CALP>
F:542-573/Domain: calmodulin repeat homology <EF1>
F:585-617/Domain: calmodulin repeat homology <EF2>

F:618-647/Domain: calmodulin repeat homology <EF3>
F:650-682/Domain: calmodulin repeat homology <EF4>
F:683-714/Domain: calmodulin repeat homology <EF5>
F:2/Modified site: acetylated amino end (Ser) (in mature form) #status experimental
F:115,272,296/Active site: Cys, His, Asn #status predicted

Query Match 16.9%; Score 618.5; DB 1; Length 714;
Best Local Similarity 32.3%; Pred. No. 1.5e-40;
Matches 166; Conservative 73; Mismatches 198; Indels 77; Gaps 17;

QY 13 LFRDAAPPAADSSL-FCDLSTPLAQFREDITWRRPQECATPRLFPDPPREGQVKGSLG 71
Db 55 LFRDEAPPPVQSLGKDLG-PNSSTYGIKKRPTLLSNPQFIVDGTATDCCGALG 113
QY 72 DCFELCAALAKSRILLDOVIPGQSPADQYRGSFTCRIMQFGRWVVTDDRLPCL 131
Db 114 DCVLLAAIASLTLDLHLLHVVPHGOS--FONGYAGIFHFQWFGWVVDVDDLLPIK 171
QY 132 AGRICFSRCQREDFVWLPILLEKVKVYAKVHSGYHLMGAVADALVDLTGGLAERWNLKGA 191
Db 172 DGLVVFHSAEGNEFWSALLEKAYAKVNGSYEALSGSTSEGFEFTGGVTEWELKAP 231
QY 192 GSGGQDQDRPGRWEHRTCRQLLHLDQCLISC-----CVLSPRA-GARELGEFHFIVSDL 245
Db 232 SDLYQ-----IILKALERSLLGCSIDISSVLDMEAITFKLVKCHAYSVTGA 279
QY 246 RELOGAGQCIILLRTONPWRCRCWGLWRGEGSGVQVDAVASSELLSOLQGEFVVEE 305
Db 280 KQV-NYRGQVSVLIRNPNWGEYETWGAUSSSSNNVDPYERDQLRVKMGDEGFWMF 338
QY 306 EEFLEFDELTVGYPVTEAGHLQSLYTERLILCHTRALPCAWYKQSGAGCGRNNGSPS-- 363
Db 339 RQPRREFTRLEICNLTPDA--LKSRTIRK--WNTTYEGTWRRGSTAGGCRN---TPATF 391
QY 364 --NPKFWLRVSEPE-----VYIAVLQSRSLHAADWAGRALVGDSTHWSWP 409
Db 392 WYVNPQFKIRLDETDDDDDYDREGSCSFVLLALMOKHR-----RRRRP----- 434
QY 410 ASIPGKHQAQVGLHLWKVKRRVNLPRV-LSMPPVAVGTACHAYD-----REVHLRCELS 462
Db 435 ---GRDMETIGFVAVYEPPELVGPVAVHLKRDFFFLANASRARSEQFINLREYSTRFLP 490
QY 463 PGYILAVPSTFELKDPAGEFLLRVF---STGRVSL 493
Db 491 PGEYVVPSTFEFHKHEADFILRVFT 524

RESULT 3
CIHUH3
N:Alternate names: calpain chain 3 - human
C:Species: Homo sapiens (man)
C:Date: 28-Apr-1995 #sequence_revision 09-Aug-1997 #text_change 16-Jul-1999
C:Accession: A56218; A34488
R:Richard, I.; Broux, O.; Allamand, V.; Fougereousse, F.; Chiannikulchai, N.; Bourg,
J.A.; Fardeau, M.; Jackson, C.E.; Cohen, D.; Beckmann, J.S.
Cell 81, 27-40, 1995
A:Title: Mutations in the proteolytic enzyme calpain 3 cause limb-girdle muscular dy:
A:Reference number: A56218; MUID:95236448
A:Accession: A56218
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-821 <RIC>
R:Cross-references: GB:X85030; NID:g791039; PIDN:CAA59403.1; PID:g791040
R:Sorimachi, H.; Imajoh-Ohmi, S.; Emori, Y.; Kawasaki, H.; Ohno, S.; Minami, Y.; Suzu
J. Biol. Chem. 264, 20106-20111, 1989
A:Title: Molecular cloning of a novel mammalian calcium-dependent protease distinct
A:Reference number: A34488; MUID:90062125
A:Accession: A34488
A:Molecule type: mRNA
A:Residues: 44-445, 'AA', 448-458, 'P', 460-461, 'P', 463-484, 'T', 486-821 <SOR>
C:Genetics:

A:Reference number: A48764; MUID:93374936

A:Accession: A48764

A:Status: preliminary

A:Molecule type: DNA; protein

A:Residues: 1-703 <SOR>

A:Cross-references: GB:DL4479; NID:9441199; PIDN:BAA03370.1; PID:9441200

A:Experimental source: stomach

A>Note: sequence extracted from NCBI backbone (NCBIN:137770, NCBIP:137771)

A:Accession: B48764

A:Status: preliminary

A:Molecule type: DNA; protein

A:Residues: 1-379, 'SS' <SO2>

A:Cross-references: GB:DL4480; NID:9441201; PIDN:BAA03371.1; PID:9495223

A:Experimental source: stomach

A>Note: sequence extracted from NCBI backbone (NCBIN:137773, NCBIP:137775)

C:Superfamily: calpain large chain; calmodulin repeat homology; calpain catalytic domain

C:Keywords: alternative splicing; cysteine proteinase; EF hand; hydrolase

F:75-327/Domain: calpain catalytic domain homology <CALP>

F:532-563/Domain: calmodulin repeat homology <EF1>

F:575-607/Domain: calmodulin repeat homology <EF2>

F:640-672/Domain: calmodulin repeat homology <EF3>

F:105,262,286/Active site: Cys, His, Asn #status predicted

Query Match 16.8%; Score 612; DB 2; Length 703;
Best Local Similarity 31.2%; Pred. No. 4.7e-40;
Matches 155; Conservative 87; Mismatches 197; Indels 58; Gaps 14;

QY 13 LFRDAAPFAADSSL-FCDLSTPLAQFREDITWRPQICATPRLFPDDPREGQVKGGLG 71
DB 45 LFKDEPPACPSALGYKDLG-PGSPDFQIGVWRKPTLCNPQFIVGATRTDIRQGLG 103
QY 72 DCWFLCACAALQKSRHLLDOVIPPQSPWADQVYSGFTCRIMQFGRWVEVTTDRPLCL 131
DB 104 DCWLLAAIAISLTNEKLLVLPVLDQS--FQKDVAGIFHFQFQYGEWVEVDDRLPTK 161
QY 132 AGRLCFSRCOREDFWFLPLEKYYAKVHGSYEHLMWAGQVADALVDLTGGLAERNLKV 191
DB 162 NGQLLHSEEGNEFWSALLEKAYAKUNGSEALVGGSTIEGTFEGFSGSEFYDLK--- 218
QY 192 GSGGQDPRGWEHRTCRQLLHLKDOCLISCCV--LSPRAGA-----RELGEFAFVSD 245
DB 219 -----KPPENLYIITQKA--LRKSGLLGCSIDVSTAAEAATRKQKLVGHAYSVTG 269
QY 246 RELQGOAGQCCILLRIQNPGRWCQGLNREGGEGSWQDAVAASELLSLQGEFVWEE 305
DB 270 EYNFVH-GRPEKLRINPGEVWSEWGSWADNAPWNYIDPRKKEELDKAEDGEFWMF 328
QY 306 EEFRLREFDELTVGYPVTEAGHLQSLYTERLLCHTRALPGAWKVGOSAGGCNNSG-FPSN 364
DB 329 SDFLKQYSRLICNLSDPSLSSEIHKWNLVFN-----GRWTRGSTAGGCLNPGTYWTN 384
QY 365 PKFWLVRVSEPE-----VYIAVLRSLRHAADWAGRARALVGDSDHTS--WSPA 410
DB 385 PQFKIHLDEVEDQEGTSEPCCTVLIGLMOKNR-----RRQKRIQOGMLSIGVAVY 436
QY 411 SIPECKHQVAGLHLKWKVEKRVNLPVLSMPVAGTACHAYDREHVLRCBLSPGYIYLV 470
DB 437 QIPELESHTDAHLGR-----DFFLGRQPSCTCSYTNLNRVSSRVRLPPGQIVLVP 488
QY 471 STFLKDPAGFEFLLRVFS 487
DB 489 STEFFPKDGFCLRVFS 505

RESULT 6
CIHUH2

N:Alternate names: calpain chain 2 - human

C:Species: Homo sapiens (man)

C:Date: 21-Nov-1993 #sequence_revision 09-Aug-1997 #text_change 16-Jul-1999

C:Accession: S10590; A31218; A33529

R:Sorimachi, H.; Ohmi, S.; Emori, Y.; Kawasaki, H.; Saido, T.C.; Ohno, S.; Minami, Y.; S

Biol. Chem. Hoppe-Seyler 371(Suppl.), 171-176, 1990
A:Title: A novel member of the calcium-dependent cysteine protease family.
A:Reference number: S10589; MUID:90380278
A:Accession: S10590
A:Molecule type: mRNA
A:Residues: 1-700 <SOR>
R:Imajoh, S.; Aoki, K.; Ohno, S.; Emori, Y.; Kawasaki, H.; Sugihara, H.; Suzuki, K.
Biochemistry 27, 8122-8128, 1988
A:Title: Molecular cloning of the cDNA for the large subunit of the high-Ca(2+)-requ
A:Reference number: A31218; MUID:89166474
A:Accession: A31218
A:Molecule type: mRNA; protein
A:Residues: 1-210, 'I', 212-394, 'D', 396-445, 'I', 447-700 <IMA>
A:Cross-references: GB:M23254; NID:9511636; PIDN:AAA35645.1; PID:9511637
A:Note: parts of this sequence were determined by protein sequencing; the amino end
R:Hata, A.; Ohno, S.; Akita, Y.; Suzuki, K.
J. Biol. Chem. 264, 6404-6411, 1989
A:Title: Tandemly reiterated negative enhancer-like elements regulate transcription
A:Reference number: A33529; MUID:89197947
A:Accession: A33529
A:Molecule type: DNA
A:Residues: 1-67, 'G', 69-72, 'IE', 75-78, 'R' <HAT>
A:Cross-references: DDBJ:J04700; NID:9179910; PIDN:AAA52760.1; PID:9463086
C:Genetics:
A:Gene: GDB:CAPN2; mCAPN; CAPNml
A:Cross-references: GDB:119750; OMIM:114230
A:Map position: Ipter-Iqter
C:Complex: heterodimer of L (large) and S (small) chains
C:Function:
A:Description: catalyzes the hydrolysis of peptides
A:Note: cleaves preferentially after tyrosine, methionine, or arginine residues and
C:Superfamily: calpain large chain; calmodulin repeat homology; calpain catalytic do
C:Keywords: acetylated amino end; calcium binding; cysteine proteinase; EF hand; het
F:2-700/Product: calpain large chain 2 #status predicted <MAT>
F:75-327/Domain: calpain catalytic domain homology <CALP>
F:529-560/Domain: calmodulin repeat homology <EF1>
F:572-604/Domain: calmodulin repeat homology <EF2>
F:605-634/Domain: calmodulin repeat homology <EF3>
F:637-669/Domain: calmodulin repeat homology <EF4>
F:2/Modified site: acetylated amino end (Ala) (in mature form) #status predicted
F:105,262,286/Active site: Cys, His, Asn #status predicted

Query Match 16.7%; Score 608; DB 1; Length 700;
Best Local Similarity 30.8%; Pred. No. 9.7e-40;
Matches 159; Conservative 74; Mismatches 209; Indels 74; Gaps 15;

QY 10 ARELFRDAAPFAADSSL-FCDLSTPLAQFREDITWRPQICATPRLFPDDPREGQVKG 68
DB 42 AGTLFODPSPPAIPSAIGFKELG-PYSSKTRGMWRKPTICADPQFIIGGATFTDICQ 100
QY 69 LLGDCWFLCACAALQKSRHLLDOVIPPQSPWADQVYSGFTCRIMQFGRWVEVTTDR 128
DB 101 ALGDCWLLAAIAISLTNEELIARVPLNQS--FOENYAGIFHFQFQYGEWVEVDDRL 158
QY 129 PCLAGRLCFSRCOREDFWFLPLEKYYAKVHGSYEHLMWAGQVADALVDLTGGLAERNL 188
DB 159 PTKDGEILLFVHSAEGSEFWSALLEKAYAKUNGSEALVGGATTEGTFEGFDTGGLAEWELK 218
QY 189 GYVAGSGQDPRGWEHRTCRQLLHLKDOCLISCCVLSPPRAGARELGEF-----HAFIV 242
DB 219 -----KPPNLFKIITQKA--LOKSGLLGCSIDITSAADSEAITFOKLKVGHAYSV 266
QY 243 SDLRELQGOAGQCCILLRIQNPGRWCQGLNREGGEGSWQDAVAASELLSLQGEFEW 302
DB 267 TGAEEVESN-GSLQKLRINPGEVWTRWVNDNCPSWNTIDPEERLTRRHEDEFEW 325
QY 303 VEDEEFLREFDELTVGYPVTEAGHLQSLYTERLLCHT-----RALPGAWKVGOSAGGCR 356
DB 326 MFSDFPLRHYSRL-----EICNLTPDTLTSDTYKKWKLTKMDGNWRRGSTRAGCR 375
QY 357 N-NSGSPSPKFWLVRVSEPE-----VYIAVLRSLRHAADWAGRARALVGDSDHT 405

Db 376 NYNTEFWNPNQYLKLEEEDEDEDSCTFLVGLQKR-----RQRKMGEDMHT 428
 QY 406 -SNSPASIRPKHQAVGLHMK-----VEKRVNLPRVLSMPPVAGTACHAVDEVLHRC 460
 Db 429 ICGEIVEVEEELSGQTNLHLSKNEFLLNRNRRSDPEINL-----REVNLRFK 476
 QY 461 LSPGYLAVSPFELKDAPEGFLIRVSTGVLSAI 496
 Db 477 LPEGEIIVPSTPEPKNDGDFCIRVSEKADYQAV 512

RESULT 7
 B34488
 calpain (EC 3.4.22.17) large chain 3 - rat

N:Alternate names: cysteine proteinase

C:Species: Rattus norvegicus (Norway rat)

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C:Accession: B34488; S10589

R:Sortmach: H.; Imajob-Ohmi, S.; Emori, Y.; Kawasaki, H.; Ohno, S.; Minami, Y.; Suzuki, J. Biol. Chem. 264, 20106-20111, 1989

A:Title: Molecular cloning of a novel mammalian calcium-dependent protease distinct from

A:Reference number: A94688; MUID:90062125

A:Accession: B34488

A:Molecule type: mRNA

A:Residues: 1-821 <SOR>

A:Cross-references: GB:J05121; NID:q205955; PIDN:AAA41790.1; PID:q205956

R:Sortmach: H.; Ohmi, S.; Emori, Y.; Kawasaki, H.; Salido, T.C.; Ohno, S.; Minami, Y.; S

Biol. Chem. Hoppe-Seyler 371(Suppl.), 171-176, 1990

A:Title: A novel member of the calcium-dependent cysteine protease family.

A:Reference number: S10589; MUID:90380278

A:Accession: S10589

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-50, 'E', '52-211, 'V', '213-252, 'K', '254-821 <SO2>

C:Superfamily: calpain large chain; calmodulin repeat homology; calpain catalytic domain

C:Keywords: calcium binding; cysteine proteinase; EF hand; hydrolase

F:99-400/Domain: calpain catalytic domain homology #status atypical <CALP>

F:649-680/Domain: calmodulin repeat homology <EF1>

F:652-724/Domain: calmodulin repeat homology <EF2>

F:725-754/Domain: calmodulin repeat homology <EF3>

F:757-789/Domain: calmodulin repeat homology <EF4>

F:790-821/Domain: calmodulin repeat homology <EF5>

F:129,334,358/Active site: Cys, His, Asn #status predicted

Query Match 16.4%; Score 599.5; DB 1; Length 821;
 Best Local Similarity 31.0%; Pred. No. 5.6e-39;
 Matches 170; Conservative 78; Mismatches 186; Indels 115; Gaps 18;

QY 13 LFRDAAPPAADSSLCFCDLSTPLAQRFDITWRPQICATPRLPPDPREGQVKGGLGD 72
 Db 74 LYPDPEPPDPTSLFYQKFPF-OF-----VWKRPPEICENRFLITIGANKRFDICQGLGD 128
 QY 73 CWFICACAAALQKSHLLDOYIPGQSPMAQDEYRGSTCRIMQFGKRVETTDRLPCLA 132
 Db 129 CWLLAATACLTLLERLFRVYIPHDQ--FTENYAGIFHFQFWRGDMVVDVYIDCLPTYN 186
 QY 133 GRLCFSRCQREDEVFMLPLEKRYAKVGSYEHLMAGQVADALVDLTGLAERMLK---- 188
 Db 187 NQVLFKRSNHRNEFWMSALLKAYAKLKGSTELAKGNTTEMEDPTGCVGEFEIKAPPS 246
 QY 189 -----GVAGSGGQDRPGRWHRFCROLDHLKDDCLI 220
 Db 247 DMTKIMKALERSGLMGCSIDDTNMTYGTSPGSLNN-----GELLARVRNN-----DMSL 299
 QY 221 SCCVLSPRAG-----ARLGEHAFIYSDRLQOAGOCILLRIQN 263
 Db 300 RQSDLDPRASDDRRPRTIYVQYETRMAGGLVGHAYSVGLEALFK--GEKVKLVLRN 358
 QY 264 PMGRRCMOGIMRGEGMSQVDAVAASELLSOL--QSGEEFVEEERFLREDELTVGIPVT 322
 Db 359 PMQGVEMNGSMGDKMDMSYVDDEKARQHOVTEDEGEFMWSTDFVYHFTKLEICMTA 418

QY 323 EAGHLQSLYTERLCHTRAL-PCAWYKGSAGGCRNNSGPF-----SNPKFWLAVSE----- 373
 Db 419 DA-----LESCKLQTVTSVNBGRWAGCSAGGCRN---FPDQFWNPNQYRLLEEDDD 470
 QY 374 --PSEV-----YIAVLQSRSLHAADWAGARALVGDSTHSWSPASIPGKHQAVGLHMKV 427
 Db 471 PDDSEVICSPFLVLMQKNR-----RKDKL-----GANLFTIGFAIYEV 509
 QY 428 -----EKRRVNLPRVLSMPVAGTACHAVDEVLHRCLESPGYLAVSPFELKDAPEG 481
 Db 510 PKEMGNKQHLQKDFFLYNASKARSKTYIMREVSRFLRPEEYIVPSTPEPHQGER 569
 QY 482 LTRVFSTGR 490
 Db 570 ILRVFSEKR 578

RESULT 8
 S57194

calpain (EC 3.4.22.17) large chain 2 - chicken

N:Alternate names: m-calpain heavy chain

C:Species: Gallus gallus (chicken)

C:Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 20-Jun-2000

C:Accession: S57194

R:Sortmach: H.; Tsukahara, T.; Okada-Ban, M.; Sugita, H.; Ishiura, S.; Suzuki, K.

Biochim. Biophys. Acta 1261, 381-393, 1995

A:Title: Identification of a third ubiquitous calpain species - chicken muscle exp

A:Reference number: S57194; MUID:95260862

A:Accession: S57194

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-700 <SOR>

A:Cross-references: EMBL:D38026; NID:q882068; PIDN:BA07228.1; PID:q882069

C:Superfamily: calpain large chain; calmodulin repeat homology; calpain catalytic d

C:Keywords: cysteine proteinase; EF hand; hydrolase

F:75-327/Domain: calpain catalytic domain homology <CALP>

F:529-560/Domain: calmodulin repeat homology <EF1>

F:572-604/Domain: calmodulin repeat homology <EF2>

F:605-634/Domain: calmodulin repeat homology <EF3>

F:637-669/Domain: calmodulin repeat homology <EF4>

F:670-700/Domain: calmodulin repeat homology <EF5>

F:105,262,286/Active site: Cys, His, Asn #status predicted

Query Match 16.4%; Score 597.5; DB 2; Length 700;
 Best Local Similarity 31.2%; Pred. No. 6.5e-39;
 Matches 163; Conservative 81; Mismatches 203; Indels 75; Gaps 17;

QY 1 MRAGRGATPARELFRDAAPPAADSSLCFCDLSTPLAQRFDITWRPQICATPRLPPDP 60
 Db 40 LQGR-----LFHDSFFPAGPALGYRELPGNSYKTKGVWCRPTLSCPPRIAGA 92
 QY 61 REGVKGGLDQWFLACALQKSHLLDOYIPGQSPMAQDEYRGSTCRIMQFGKRVETTDRLPCLA 120
 Db 93 TRDIDCGALGDDCWLLAATACLTLLERLFRVYIPHDQ--FTENYAGIFHFQFWRGDMVVDVYIDCLPTYN 186
 QY 121 EYTTDRRLPCLAGRLCFSRCQREDEVFMLPLEKRYAKVGSYEHLMAGQVADALVDLTGL 180
 Db 151 DVVVYDDRLPTKNGELLFVHSAESESFALLERKAYAKLNGSYELSGTTTEGEDEFDGG 210
 QY 181 LAERMLNKGVAAGSGGQDRPGRWHRFCROLDHLKDDCLI 220
 Db 211 IAEYTEL-----QKAPNLF-----KIQRALQKSLGCSIDTISAETEAVTSQ 256
 QY 233 ELGEHAFIYSDRLQOAGOCILLRIQNPMGRRCMOGIMRGEGMSQVDAVAASEL 292
 Db 257 KIVKGAHYSTGABEVRN-R-GSIOKLRIRNPMGEVEMTKMDNCNMGCVPEVNERL 315
 QY 293 LSQLOEGEFVEEERFLREDELTVG--YPVTEAGHLSQSLYTERLCHTRALPGAWYKQ 350
 Db 316 TRRHEGGEFMAFNNDPLRHYSRLICNLFPDILASDYKTKMS--LL-----KLGGNMRGA 369
 QY 351 SAGGCRN-NSGFPNSRKFMLRVSEPS-----YIAVLQSRSLHAADWAGARAL 399

Db 370 TAGGCRNPTNFWNPOLIKLEEDDDPPDEGGCTFLGLIQKHR-----RKORKM 422
Qy 400 VGDST-SWSPASIPGKHYQAVGLRWK-----VEKRVNLPVLSMPVAGTACHAYDRE 454
Db 423 GEDMTIGFAIYVPPFSGTNIHLSKNFPLTNKREKSNFTNL-----RE 470
Qy 455 VHLRCELSPGYLLAVPTFLKADPGEFLLRVFTGRVLSAI 496
Db 471 VLNRFKLPAAGEYIIVPSTFENLNGDFCLRVSEKNNSTVI 512

RESULT 9
S57196
calpain (EC 3.4.22.17) large chain 3 - chicken
N:Alternate names: calpain p94 heavy chain; n-calpain-1 heavy chain
C:Species: Gallus gallus (Chicken)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
A:Accession: S57196
R:Sorinachi, H.; Tsukahara, T.; Okada-Ban, M.; Sugita, H.; Ishiura, S.; Suzuki, K.
Biochim. Biophys. Acta 1261, 381-393, 1995
A:Title: Identification of a third ubiquitous calpain species - chicken muscle expressed
A:Reference number: S57194; MUID:95260862
A:Accession: S57196
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-810 <SOR>
C:Cross-references: EMBL:D38028; NID:g882072; PIDN:BA07230.1; PID:g1552167
C:Superfamily: calpain large chain; calmodulin repeat homology; calpain catalytic domain
C:Keywords: cysteine proteinase; EF hand; hydrolase
F:93-393/Domain: calpain catalytic domain homology <CALP>
F:746-778/Domain: calmodulin repeat homology <EFH>

Query Match 16.2%; Score 589.5; DB 1; Length 810;
Best Local Similarity 28.6%; Pred. No. 3.3e-38;
Matches 161; Conservative 87; Mismatches 178; Indels 137; Gaps 19;
Qy 13 LFRDAAPADSSLCFCDLSTPLAQFEDITWRPQECATPRLPDPDPREGOVKQGLIGD 72
Db 68 LYEDDPFPPTNLSFYKQVPI-----KFEWKRPREICENPRFTIGGANTDQCQGLGD 122
Qy 73 CWFCAACALQKSHLLDQVIPPQPSWADQEVGSGTCRIWQFGRVVEVTTDDRLPCLA 132
Db 123 CWFALATCLTNKLLCRVPHDQS---FTQYAGIFHFQFQWYGVWDVYIIDDCPLTYN 180
Qy 133 GLCFSCRCOREDFVWLPLEKVKYAKVHGSYEHVWAGQVADALVDLTGLGLAERNLK 188
Db 181 NLVFTKSSQNEEWSALLEKAKYALHGSYALKGGNTTAMEDFTGCVTEFTKIDAPK 240
Qy 189 -----GVAGSG----- 194
Db 241 DIYKIMKHAIARGSLMASSIDDLGPHYGAAPRSDIGELIARMVKLENNAQMTYSTVDYQ 300
Qy 195 GOODRPGRWEHRTROLLHLKQCLISCCVLSPPAGARELGEHAFIVSDRLRQAGQAG 254
Db 301 GTERPA-W-----TIMPQVETRMSC-----GLVKHAYSVTAVEETT-YKGE 342
Qy 255 CILLRIQNPWGRRCWGLRREGGWSOVDAVAASELLSOL-QEGEFVVEEEFLREED 313
Db 343 KMLVRLRNPNWQVWNGPNSDSEENFIDEEEKIRLQHKIAEDGEFVWISLEDFMRHET 402
Qy 314 ELTVGYPVTAGHLQSLYTERLCHTRAL-PGAWRVQSGAGGCRN-NSGFPSPNPKFWLRV 371
Db 403 KLEICNLTPD-----TLEADKLQTVSVNVEGRVRCSSAGGCRNPDFTWNPYRLKL 457
Qy 372 SEPE-----VYIAVQSRHLHADWAGARALVGDST-SWSPASIP-----GK 415
Db 458 LEEDDDPEDEVICSLFVALMQNR-----RKRIKANLYTIGFAIYEVKPMKHGK 510
Qy 416 HYQAVGLHWKVEKR-----VNLPRVLSMPVAGTACHAYDREHVRCELSPGYLLAVPS 471
Db 511 HHLQKDFELNASKARSKTYIN-----REISERFLRPSEYVIPS 552

Qy 472 TFLKADPGEFLLRVFTGRVLS 494
Db 553 TYEPHQEGEFLLRVFESEK-SLS 574
RESULT 10
A55054
calpain (EC 3.4.22.17) large chain - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 08-Jul-1995 #sequence_revision 03-Aug-1995 #text_change 18-Jun-1999
C:Accession: A55054
R:Emori, Y.; Saigo, K.
J. Biol. Chem. 269, 25137-25142, 1994
A:Title: Calpain localization changes in coordination with actin-related cytoskeleton
A:Reference number: A55054; MUID:95014293
A:Accession: A55054
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-805 <EMO>
A:Cross-references: GB:X78555; NID:g562287; PIDN:CAA55297.1; PID:g562288
C:Genetics:
A:Gene: FlyBase:Calpa
A:Cross-references: FlyBase:FBgn0012051
C:Superfamily: calpain large chain; calmodulin repeat homology; calpain catalytic do
C:Keywords: cysteine proteinase; duplication; EF hand; hydrolase
F:50-347/Domain: calpain catalytic domain homology <CALP>
F:576-708/Domain: calmodulin repeat homology <EF1>
F:709-738/Domain: calmodulin repeat homology <EF2>
F:741-773/Domain: calmodulin repeat homology <EF3>
F:774-805/Domain: calmodulin repeat homology <EF4>

Query Match 16.1%; Score 588; DB 1; Length 805;
Best Local Similarity 30.0%; Pred. No. 4.4e-38;
Matches 162; Conservative 84; Mismatches 200; Indels 94; Gaps 18;
Qy 13 LFRDAAPADSSLCFCDLSTPLAQF--RED--ITWRPQECATPRLPDPDPREGOVKQ 68
Db 65 LFEDPLFPASNESL-----QFSRPDRHLEWLPHEIAENPQFVGVSEDFVQOG 115
Qy 69 LLGDCWFLCACAALQKSHLLDQVIPPQPSWADQEVGSGTCRIWQFGRVVEVTTDDL 128
Db 116 ELGDCWLLATANTLQESNLFRRVPAEQS--FEENYAGIFHFQFQWYGVWDVYIIDRL 173
Qy 129 PCLAGLRCSCOREDFVWLPLEKVKYAKVHGSYEHVWAGQVADALVDLTGLGLAERNLK 188
Db 174 PTYNGELMYHMHSTKNEFWSALLEKAKYALHGSYALKGGSTCEAMEDFTCGVSEWYDLK 233
Qy 189 GVAGS-----GGQDRPGRWEHRTROLLHLKQCLISCCVLSPPAGAR 232
Db 234 EAPGNLTILQKAERNMGMCSIEPDNPVTEAETPGILRGHAYSITKVCILID----- 287
Qy 233 ELGEFHAFIVSDRLRQAGQCLILLRIQNPWGRRC-WOGLMREGGWSQVDAVAASE 291
Db 288 -----IVTPNR--QK-----IPMIRNPNWGNENWPNWSDSPENRYIPEQKAE 333
Qy 292 L-LSOLQEGEFVVEEEFLREDFDELTVGYVPTAGHLQSLYTERLCHTRALPGAWVKG 350
Db 334 IGLTDRDGEFWMSQDFLNLHFDVREICNLSPDSLTDQHQHGRKRWKEMSYEGEWTPGV 393
Qy 351 SAGGCRN-NSGFPSPNPKFWLRVSEPE-----VYIAVQSRHLHADWAGARALVG 401
Db 394 TAGGCRNFLTDFHNNQYIITVDPDEDEEGCTVIVALMQNR-----RSKRNM- 444
Qy 402 DSHTSWSPASIPGKHYQAVGLHWKVEKR--NLPRVLSM---PPVAGTACHAYD-REH 456
Db 445 -----GMECLTIGFAIYSLNDRLENRPNQGLNFRYKSSVGRSPHINTREVC 492
Qy 457 LRCELSPGYLLAVPTFLKADPGEFLLRVFTGRVLS-----AIRAVAKNTTPGAALP 510
Db 493 ARFKLPGGHYLIIVPSTFDPNEEGEFTIRVSETQNNMEENDHVGYGKADTTTPGFPPT 552

RESULT 11

S44749
C06G4.2 protein - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S44749
R:Waterston, R.
submitted to the EMBL Data Library, November 1993

A:Description: Sequence of the C. elegans cosmid C06G4.
A:Reference number: S44747
A:Accession: S44749
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-653 <WAT>
A:Cross-references: EMBL:L25598; NID:g409290; PIDN:AAA27940.1; PID:g409293
C:Genetics:
A:Introns: 26/1; 216/3; 269/3; 357/2; 401/2; 558/1; 583/3; 613/3
C:Superfamily: C06G4.2 protein; calpain catalytic domain homology
F:245-498/Domain: calpain catalytic domain homology <CALP>

Query Match 15.9%; Score 579; DB 1; Length 653;
Best Local Similarity 30.8%; Pred. No. 1.7e-37;

Matches 157; Conservative 69; Mismatches 176; Indels 108; Gaps 16;

```

OY 10 ARELFRDAFPAADSSLRCDLSTPLAOPREDITWRPRPOICATPRLEFPDDPREGOVKGL 69
DB 217 SRLLEDPOFLANDSLEFSKRP-----KREWLPRGELITREPOLITEGHSRFDYIOGE 271
OY 70 LGDCWFLCACAALQKSRHLDDVIPPQOPSMADQERSGFTRIQOFGRWVEYTTDDRLP 129
DB 272 LGDCWFLCACAALQKSRHLDDVIPPQOPSMADQERSGFTRIQOFGRWVEYTTDDRLP 129
OY 130 CLAGRLCFSRCQREDVFWLPLEKRYAVKVSYEHLNAGOVADALVDLGLAERWNLK 188
DB 330 TSNGLTLYMHSANNEFSALEKAYAKLFSGSYEALKGTTSEALQEDMKGLEFIDLKN 389
OY 189 -----GVAAGSGGODRGRGWRHRTCRQLHLKQCLISCCVLSPPAGAR 232
DB 390 PPRNIMOMMAGFENGSLFGCSIEAD-FNVMWAKMSNGELY----- 428
OY 233 ELGEHFIVSDRLAQOAGQ-CILLRIONPMG-RRCOMGLMREGGSGQVDAAVAS 290
DB 429 ---KGHAISTIGCRIVDPNGOTCT-LRIRNPMGEBOMGPMWSDNSREMSVSDYKQ 483
OY 291 EL-LSQLQGEFWEVEEFLREDELTV--GYPV-TEAGHLQSLYTERLCHTRALPGA 345
DB 484 DMGLKTFHDGGEFMSFDDFMNFEKMEICNLGPDVADRYQMTGVKAGMVAANTHGA 543
OY 346 WVKQAGGAGCGNN-SGEPSPKFWLRY--SEYIYAVLQSRHLAADMAGRARAL 399
DB 544 WVKQAGGAGCGNNYINTFANNPOFRVQLTSDPDDELCTAGNNNGRLSKOFFAANKSAM 603
OY 400 VGDSTHSWSPASIPGKHQYAVGLHLMKYEKRRVNLPRVLSMPVAGTCAHADRPHLC 459
DB 604 RSAAF-----INL-----RMKTRGF 618
OY 460 ELSPGYLAVSTFLKADAPGELLRVSTG 489
DB 619 RVPGNVYVVPSTFEPNEAEFMLRVYTN 648

```

RESULT 12

JC5772

tissue-specific calpain htra-3 - human

C:Species: Homo sapiens (man)

C:Date: 24-Jan-1998 #sequence_revision 13-Mar-1998 #text_change 18-Jun-1999

C:Accession: JC5772

R:Mugita, N.; Kimura, Y.; Ogawa, M.; Sawa, H.; Nakao, M.

Biochem. Biophys. Res. Commun. 239, 845-850, 1997

A:Title: Identification of a novel, tissue-specific calpain htra-3; a human homologue of

A:Reference number: JC5772; MUID:98042481

A:Accession: JC5772

A:Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-647 <MUG>

A:Cross-references: GB:U94346

C:Comment: This protein acts as a cellular receptor of calcium ions in physiologic

C:Genetics:

A:Gene: htra-3

A:Map position: 11q14

C:Superfamily: human tissue-specific calpain htra-3; calpain catalytic domain homo

F:1-48/Domain: I #status predicted <IND>

F:49-309/Domain: cysteine protease #status predicted <CPT>

F:31-329/Domain: calpain catalytic domain homology <CALP>

F:310-505/Domain: III #status predicted <III>

F:317-334/Domain: calcium binding #status predicted <CAL>

F:506-639/Domain: T #status predicted <TDD>

Query Match 15.6%; Score 569.5; DB 2; Length 647;
Best Local Similarity 30.9%; Pred. No. 9.4e-37;

Matches 159; Conservative 76; Mismatches 196; Indels 83; Gaps 18;

```

OY 13 LFRDAFPAADSSLRCDLSTPLAOPREDITWRPRPOICATPRLEFPDDPREGOVKGLD 72
DB 26 LREDPLEFATDLSLYK-GTPGPAVR---WKPRKICDPRLFVDGISSHDLHGOVGN 80
OY 73 CWFLLCACAALQKSRHLDDVIPPQOPSMADQERSGFTRIQOFGRWVEYTT 124
DB 81 CWFVAACSLASRESLMQVY----PDWKEQEDPRKQAYAGIFHFHWRLG-MVDVVI 135
OY 125 DRLPLCAGRLCFSRCQREDVFWLPLEKRYAVKVSYEHLNAGOVADALVDLGLAER 184
DB 136 DERLPLTYVQNLQYCHSNSNEFEKALYERAKYAKIAGCYALDGVNTDALVDFTGVSER 195
OY 185 WNLKGVAGSGGODRGRGWRHRTCRQLHLKQCLISCCVLSPPAGARL-----G 235
DB 196 IDLREGDFRNDTTRKNQLE----RLKAVHSRGGLISIKAKVTADMARLACGLYK 251
OY 236 EHFHFIVSDRLAQOQ-----GQCILLRIONPMGRRCOMGLMREGGSGQVDA 287
DB 252 THAVAVTDVRYKRVLLTHLTLAFKSEKIDMLRLNPMGRMNPMSPTSEMKVSKS 311
OY 288 VASELISQLQ-EGEWEVEEFLREDELTVGYPVTEAGHLQSLYTERLCH--TRAL 342
DB 312 EREKGVTVODDGEFFMNGEDVCR-----YEFTDIICRVINTSHLSIKTWEARL 363
OY 343 PGAVY-----KGQAGGCGNN-NSGEPSPKFWLRYSE-SEYIYAVLQSRHLAADMAGR 393
DB 364 HCAWTLHEDPRKNGGCGCINHKOTGF-QNPQYIFEVKRPEDVLLCICDQRPFRSTRERE 422
OY 394 GRARLVGDSTHSWSPASIPGKHQYAVGLHLMKYEKRRVNLPRVLSMPVAGTCAHADR 453
DB 423 GGENUL-----AIGPDIYKVENNOY--RMSIHKAKASSIYINSR 461
OY 454 EYHLCESLSPGYLAVSTFLKADAPGELLRVSTG 487
DB 462 SVFLRTDQEGRYVITPTFEPGHGEEFLRVFT 495

```

RESULT 13

S71885

sex-determining protein transformer-3 - *Caenorhabditis elegans*

N:Alternate names: tra-3 protein

C:Species: *Caenorhabditis elegans*

C:Date: 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 29-Oct-1999

C:Accession: S71885; T23631

R:Barnes, T.M.; Hodgkin, J.

EMBO J. 15, 4477-4484, 1996

A:Title: The tra-3 sex determination gene of *Caenorhabditis elegans* encodes a membe

A:Reference number: S71885; MUID:97042339

A:Accession: S71885

A:Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-648 <BAR>
A:Cross-references: EMBL:U12920
R:Matthews, L.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19775
A:Accession: T23631
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-648 <ML>
A:Cross-references: EMBL:Z82277; PIDN:CAB05248.1; GSPDB:GN00022; CESP:LLC1.1
A:Experimental source: clone LLC1
C:Genetics:
A:Gene: tra-3
A:Map position: IV; 4
A:Introns: 78/3; 142/3; 205/3; 273/3; 394/3; 513/3; 581/3
C:Function:
A:Description: required for the correct sexual development of soma and germ
A:Note: fully dispensable in males
C:Superfamily: human tissue-specific calpain htra-3; calpain catalytic domain
F:53-336/Domain: calpain catalytic domain homology <CALP>

Query Match 15.6%; Score 569; DB 2; Length 648;
Best Local Similarity 28.8%; Pred. No. 1e-36;
Matches 168; Conservative 93; Mismatches 217; Indels 106; Gaps 2

QY 11 RELFRDAFFPAARUSSLCDSLTPLAQPREDITWRRPQEICATPLFPDDPRGGVKGGLL 70
Db : : | | : : | | : : | | : : | | : : | | : : | | : : | | : : | |
26 KQPFVDTLFPPTNOSELE-----QROSSDIVMKRGELHPDLFLVEGASPNVTQGIL 80
QY 71 GDCWFLCACAAALOKSRHLLDQVPPGPSPWADPE-----YRGSTCRIMQFGRWVEVTT 124
Db : : | | : : | | : : | | : : | | : : | | : : | | : : | | : : | |
81 GNCWFVSACSALTWNFKLLAQVIPDAD----DQEWSTKHAYAGIFRFRRFWKGEVVI 136
QY 125 DRLPCLAGRLCTSRCQREDFWMLPLEKYAKVHGYSIEHLWAGQVADALVLTGGIAER 184
Db DDLTPTDRGKLFLARSFTPNFEWFSALLEKAFKLYGCYNLVGSHLSDALQDVSGGVAET 196
QY 185 WNLKGVAGSGQQDRPCRWEHRTCOLLHLKDCLISCVCVLSPRAGARELGEP----- 237
Db : : : : | | : : | | : : | | : : | | : : | | : : | | : : | | : : | |
197 LHVRKFL-----KDDPNDETCLKLFNDUKTADFAGLVAVVAALAART--KEEIESLDGSLVK 250
QY 238 -HAFTVS-----DLRELQAGQCIL-----LLRIQNPGWRCRCQGLWRGEGEQWSQ 283
Db : : | | : : | | : : | | : : | | : : | | : : | | : : | | : : | |
251 GHAYAVSAVCTIDVTNPENRSFTSFIMGSKRKKNLIQLQNPWGCEKWNGAWSDDSPWQON 310
QY 284 VDAIVASEL-----LSOLOGEFEWEEEFRLREDELTGVPV--TEAGHLQSLYTERLIC 337
Db : : | | : : | | : : | | : : | | : : | | : : | | : : | | : : | |
311 VSASQLSTMGVQVPANSDSDGDGFMPWESVHVHTDISLCOLENTSVFSRSTDEQIVF 370
QY 338 HTRALPG--AMWRQGSAGGCCRN--NSGFPSPKFWRVLRSEPS-EVYIATVLQRSRLHADWA 393
Db : : | | : : | | : : | | : : | | : : | | : : | | : : | | : : | |
371 SEWTNCKKSGAPDDRAGGCHNFKAFCNNPVIEDIPSPNCVMFALIQNDP----- 423
QY 394 GRARALVGDSHTSWSPASIPCKHYQAVGLHLWKVEKRNVMLPRV-LSMPPVGTACHAYD 452
Db : : | | : : | | : : | | : : | | : : | | : : | | : : | | : : | |
424 -----SEGKKREFVTVMGHVMKVENNROY--RVHTAMHIPA-ISDIYASG 466
QY 453 REYHLRCELSP-GYILAAPPSTFLKDPAGE---FLLRVFTSGRYVLSAI----- 496
Db : : | | : : | | : : | | : : | | : : | | : : | | : : | | : : | |
467 RSVYLHLQSLPRGRYLLIPTTF---APKEQTFLWLRYVSDHEIHFSPLTKHAPKLGILLKC 523
QY 497 ---RAVANKVTPGAALPAGEWTVQL-----RGSNRVVGQTAG 530
Db : : : : | | : : | | : : | | : : | | : : | | : : | | : : | | : : | |
524 KSAQSIVRLTIHGVDFNNSASTGFHNHYAILKDSRKSFRTKTLSSG 567

C;Accession: A39343; A45642
 A;Andresen, K.; Tom, T.D.; Strand, M.
 J. Biol. Chem. 266, 15085-15090, 1991
 A;Title: Characterization of cDNA clones encoding a novel calcium-activated neutral
 A;Reference number: A39343; MUID:91332027
 A;Accession: A39343
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-758 <AND>
 A;Cross-references: GB:M67499; NID:gl60936; PIDN:AAA29858.1; PID:gl60937
 R;Karcz, S.R.; Podesta, R.B.; Siddiqui, A.A.; Dekaban, G.A.; Strejan, G.H.; Clarke,
 Mol. Biochem. Parasitol. 49, 333-336, 1991
 A;Title: Molecular cloning and sequence analysis of a calcium-activated neutral prot
 A;Reference number: A45642; MUID:92131071
 A;Accession: A45642
 A;Status: preliminary; not compared with conceptual translation
 A;Molecule type: mRNA
 A;Residues: 1-14, 'A', 16-120, 'A', 121-127, 'H', 129-213, 'R', 215-384, 'VTC', 388-440, 'S', 44
 A;Cross-references: GB:W74233; NID:gl60934
 A;Note: sequence extracted from NCBI backbone (NCBTP:79194)
 C;Superfamily: calpain large chain; calmodulin repeat homology; calpain catalytic dc
 C;Keywords: calcium binding; cysteine proteinase; duplication; EF hand; hydrolase
 F;124-380/Domain: calpain catalytic domain homology <CALP>
 F;628-660/Domain: calmodulin repeat homology <EF1>
 F;661-690/Domain: calmodulin repeat homology <EF2>
 F;693-726/Domain: calmodulin repeat homology <EF3>
 F;727-758/Domain: calmodulin repeat homology <EF4>
 F;154,313,337/Active site: Cys, His, Asn #status predicted
 Query Match 14.8%; Score 540; DB 1; Length 758;
 Best Local Similarity 30.0%; Pred. No. 2.4e-34;
 Matches 15; Conservative 77; Mismatches 219; Indels 66; Gaps 19;

DB 534 PSTFDPNIEVNFILRVFS--OTSITEOELEDNTNOG 568

Thu Sep 19 15:10:39 2002

us-09-768-877-2.sepl9.ra

Page 1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 19, 2002, 14:48:51 ; Search time 14.68 Seconds
(without alignments)
1118.120 Million cell updates/sec

Title: US-09-768-877-2

Perfect score: 3649

Sequence: 1 MRAGGATPARELFDAFP.....SQEMLGFLQEVSMVMVMT 672

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents.AA.*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PTCUTS.COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/Backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3649	100.0	672	4	US-09-422-869-2
2	2963	81.2	666	4	US-09-422-869-18
3	2715.5	74.4	517	4	US-09-422-869-6
4	2701.5	74.0	544	4	US-09-422-869-4
5	2699	74.0	513	4	US-09-422-869-8
6	2336	64.0	444	4	US-09-422-869-10
7	858.5	23.5	274	4	US-09-422-869-12
8	635	17.4	703	2	US-08-835-099A-1
9	635	17.4	703	2	US-09-157-349-1
10	635	17.4	712	2	US-08-835-099A-2
11	635	17.4	712	3	US-09-157-349-2
12	618.5	16.9	714	4	US-09-422-869-22
13	618	16.9	821	4	US-09-422-869-24
14	612	16.8	703	4	US-09-422-869-27
15	604	16.6	700	1	US-08-726-525-7
16	604	16.6	700	2	US-08-487-942-7
17	604	16.6	700	2	US-08-726-036A-7
18	604	16.6	700	4	US-09-422-869-23
19	604	16.6	700	4	US-09-083-516-7
20	584	16.0	639	4	US-09-422-869-25
21	561.5	15.4	639	4	US-09-422-869-28
22	513.5	14.1	139	4	US-09-422-869-14
23	483.5	13.3	641	4	US-09-422-869-26
24	472	12.9	138	4	US-09-422-869-16
25	271	7.4	854	2	US-08-928-692-17
26	264.5	7.2	842	2	US-08-928-692-18
27	133	3.6	67	3	US-09-120-365-97

28	133	3.6	67	4	US-09-515-039-97	Sequence 97, Appl
29	124	3.4	67	3	US-09-120-365-95	Sequence 95, Appl
30	124	3.4	67	4	US-09-515-039-95	Sequence 95, Appl
31	121	3.3	67	3	US-09-120-365-98	Sequence 98, Appl
32	121	3.3	67	3	US-09-515-039-98	Sequence 98, Appl
33	117.5	3.2	67	3	US-09-120-365-99	Sequence 99, Appl
34	117.5	3.2	67	4	US-09-515-039-99	Sequence 99, Appl
35	117	3.2	67	3	US-09-120-365-96	Sequence 96, Appl
36	117	3.2	67	3	US-09-515-039-96	Sequence 96, Appl
37	115	3.2	67	4	US-09-120-365-94	Sequence 94, Appl
38	115	3.2	67	3	US-09-515-039-94	Sequence 94, Appl
39	110	3.0	3519	4	US-09-428-517-4	Sequence 94, Appl
40	105	2.9	1996	2	US-08-804-198-3	Sequence 94, Appl
41	105	2.9	1996	2	US-08-804-198-3	Sequence 94, Appl
42	102.5	2.8	512	4	US-09-356-818A-2	Sequence 3, Appl1
43	102.5	2.8	548	1	US-08-247-902A-2	Sequence 2, Appl1
44	102.5	2.8	687	5	PCR-US91-09784-2	Sequence 2, Appl1
45	100.5	2.8	505	1	US-08-220-603A-10	Sequence 10, Appl1

ALIGNMENTS

RESULT 1									
US-09-422-869-2									
Sequence 2, Application US/09422869									
Patent No. 6235481									
GENERAL INFORMATION:									
APPLICANT: POLONSKY, KENNETH S.									
APPLICANT: HORIKAWA, YUKIO									
APPLICANT: ODA, NAOMISA									
APPLICANT: COX, NANCY J.									
APPLICANT: SREENAN, SEAMUS									
APPLICANT: ZHOU, YUN-PING									
APPLICANT: OTANI, KENICHI									
APPLICANT: HANIS, CRAIG L.									
TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES									
FILE REFERENCE: ARCD:307									
CURRENT APPLICATION NUMBER: US/09/422,869									
CURRENT FILING DATE: 1999-10-21									
EARLIER APPLICATION NUMBER: 60/134,175									
EARLIER FILING DATE: 1999-05-13									
NUMBER OF SEQ ID NOS: 30									
SOFTWARE: PatentIn Ver. 2.0									
SEQ ID NO 2									
LENGTH: 672									
TYPE: PRT									
ORGANISM: Human									
US-09-422-869-2									
Query Match									
Best Local Similarity 100.0%; Score 3649; DB 4; Length 672;									
Matches 672; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	MRAGGATPARELFDAFPADSSLPFLDLSPLPLQFEDITWRBPQICATPRFPDP	60						
DB	1	MRAGGATPARELFDAFPADSSLPFLDLSPLPLQFEDITWRBPQICATPRFPDP	60						
QY	61	REGGVKGLDLCWFLCACAALOKSRHLLDYTPPGQSWMDQEVRSFCTRIWFGHW	120						
DB	61	REGGVKGLDLCWFLCACAALOKSRHLLDYTPPGQSWMDQEVRSFCTRIWFGHW	120						
QY	121	EVTDDRRPCLAGRCFCRCQREDFWPLPLEKYAKYAGSYEHLMAQOVADALVDLGG	180						
DB	121	EVTDDRRPCLAGRCFCRCQREDFWPLPLEKYAKYAGSYEHLMAQOVADALVDLGG	180						
QY	181	LAERNNLKGAVSGGQDPRGWEHRCROLHLTKDCLSCVLSPPAGAREGEFNAF	240						
DB	181	LAERNNLKGAVSGGQDPRGWEHRCROLHLTKDCLSCVLSPPAGAREGEFNAF	240						
QY	241	IVSDLRLEIAGAGACILLIRIIONPWGRRCWGLMREGGEGMSQVDAVAASELSQLDGE	300						

Db 241 IVSDRLRQAGOCILLRLRIONPWRRCWOGLWREGGEGSWQVDAVAASELLSOLQEGE 300
Qy 301 FWVEEEFLREFDELTVGYPVTEAGHLQSLYTERLLCHTRALPGAWVQSGAGGCRNNSG 360
Db 301 FWVEEEFLREFDELTVGYPVTEAGHLQSLYTERLLCHTRALPGAWVQSGAGGCRNNSG 360
Qy 361 FPSNPKFWLRVSESEVYIAVQSRRLHAADWAGRARALVGDSTHSWSPASIPGKHQYAV 420
Db 361 FPSNPKFWLRVSESEVYIAVQSRRLHAADWAGRARALVGDSTHSWSPASIPGKHQYAV 420
Qy 421 GLHLWKVKRRVNLPRVLSMPVAGTACHAYDREHLCESLSPGYLLAVPSTFLKADPGE 480
Db 421 GLHLWKVKRRVNLPRVLSMPVAGTACHAYDREHLCESLSPGYLLAVPSTFLKADPGE 480
Qy 481 FLLRVFSTGRVSLSAIRAVAKNTTPGAALPAGEMGTVOLGRSWRVGQTAGGSRNFPASYPT 540
Db 481 FLLRVFSTGRVSLSAIRAVAKNTTPGAALPAGEMGTVOLGRSWRVGQTAGGSRNFPASYPT 540
Qy 541 NPCFPFSVPEGPGRCVRITLHQCRRSDTEFHPHIGFHFQVPEGGRSQDAPPLLLQBP 600
Db 541 NPCFPFSVPEGPGRCVRITLHQCRRSDTEFHPHIGFHFQVPEGGRSQDAPPLLLQBP 600
Qy 601 LSCVPHRYAQVSRCLLPAGTYKVPSTYLPDTEGAFVTIATRIDRPSIHSQEMLGQF 660
Db 601 LSCVPHRYAQVSRCLLPAGTYKVPSTYLPDTEGAFVTIATRIDRPSIHSQEMLGQF 660
Qy 661 LQEVSMVAVMKT 672
Db 661 LQEVSMVAVMKT 672

RESULT 2
US-09-422-869-18
; Sequence 18, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: ODA, NAOHISA
; APPLICANT: COX, NANCY J.
; APPLICANT: SREENAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: OTANI, KENICHI
; APPLICANT: HANIS, CRAIG L.
; APPLICANT: BELL, GRAEME I.
; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
; FILE REFERENCE: ARCD:307
; CURRENT APPLICATION NUMBER: US/09/422,869
; CURRENT FILING DATE: 1999-10-21
; EARLIER APPLICATION NUMBER: 60/134,175
; EARLIER FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 666
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-422-869-18

Query Match 81.2%; Score 2963; DB 4; Length 666;
Best Local Similarity 81.8%; Pred. No. 3.7e-289;
Matches 544; Conservative 45; Mismatches 76; Indels 6; Gaps 2;
Qy 1 MRAGRGATPARELFRDAAPFAADSSILFCDLSTPLAOPREDITWRPQECATPPLPDDP 60
Db 1 MRVRAETPARELFRDAAPFAADSSILFYNLSTPLAOPREDITWRPQECATPPLPDDP 60
Qy 61 REGOVKGLGDCWFLCACAALQKSRHLLDQVIPPQGPSWADQYGRSFTCRWQFGRW 120
Db 61 WEGOVKGLGDCWFLCACAALQKSHLLDQVIPPQGPSWADQYGRSFTCRWQFGRW 120

Qy 121 EYTTDRPLCLAGRLCFSRCQREDVFWLPLEKVKYAKVHGSYERHLWAGOVADALYDLTGG 180
Db 121 EYTTDRPLCLAGRLCFSRCQREDVFWLPLEKVKYAKVHGSYERHLWAGOVADALYDLTGG 180
Qy 181 LAERNLKGAGSGGQDRPGWEHRTCRQLHLKDCQLISCCVLSPRAGARELGEFRAF 240
Db 181 LAERNLKGAGSGGQDRPGWEHRTCRQLHLKDCQLISCCVLSPRAGARELGEFRAF 240
Qy 241 IVSDRLRQAGOCILLRLRIONPWRRCWOGLWREGGEGSWQVDAVAASELLSOLQEGE 300
Db 241 IVSDRLRQAGOCILLRLRIONPWRRCWOGLWREGGEGSWQVDAVAASELLSOLQEGE 300
Qy 301 FWVEEEFLREFDELTVGYPVTEAGHLQSLYTERLLCHTRALPGAWVQSGAGGCRNNSG 360
Db 301 FWVEEEFLREFDELTVGYPVTEAGHLQSLYTERLLCHTRALPGAWVQSGAGGCRNNSG 360
Qy 361 FPSNPKFWLRVSESEVYIAVQSRRLHAADWAGRARALVGDSTHSWSPASIPGKHQYAV 420
Db 361 FPSNPKFWLRVSESEVYIAVQSRRLHAADWAGRARALVGDSTHSWSPASIPGKHQYAV 420
Qy 421 GLHLWKVKRRVNLPRVLSMPVAGTACHAYDREHLCESLSPGYLLAVPSTFLKADPGE 480
Db 421 GLHLWKVKRRVNLPRVLSMPVAGTACHAYDREHLCESLSPGYLLAVPSTFLKADPGE 480
Qy 481 FLLRVFSTGRVSLSAIRAVAKNTTPGAALPAGEMGTVOLGRSWRVGQTAGGSRNFPASYPT 540
Db 481 FLLRVFSTGRVSLSAIRAVAKNTTPGAALPAGEMGTVOLGRSWRVGQTAGGSRNFPASYPT 540
Qy 541 NPCFPFSVPEGPGRCVRITLHQCRRSDTEFHPHIGFHFQVPEGGRSQDAPPLLLQBP 600
Db 541 NPCFPFSVPEGPGRCVRITLHQCRRSDTEFHPHIGFHFQVPEGGRSQDAPPLLLQBP 600
Qy 601 LSCVPHRYAQVSRCLLPAGTYKVPSTYLPDTEGAFVTIATRIDRPSIHSQEMLGQF 660
Db 601 LSCVPHRYAQVSRCLLPAGTYKVPSTYLPDTEGAFVTIATRIDRPSIHSQEMLGQF 660
Qy 661 LQEVSMVAVMKT 671
Db 661 LQEVSMVAVMKT 665

RESULT 3
US-09-422-869-6
; Sequence 6, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: ODA, NAOHISA
; APPLICANT: COX, NANCY J.
; APPLICANT: SREENAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: OTANI, KENICHI
; APPLICANT: HANIS, CRAIG L.
; APPLICANT: BELL, GRAEME I.
; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
; FILE REFERENCE: ARCD:307
; CURRENT APPLICATION NUMBER: US/09/422,869
; CURRENT FILING DATE: 1999-10-21
; EARLIER APPLICATION NUMBER: 60/134,175
; EARLIER FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 517
; TYPE: PRT
; ORGANISM: Human
US-09-422-869-6

Query Match 74.4%; Score 2715.5; DB 4; Length 517;
Best Local Similarity 76.8%; Pred. No. 2e-264;
Matches 516; Conservative 1; Mismatches 0; Indels 155; Gaps 1;

```

QY 1 MRAGGATPARELFRDAAPPAADSLFCDSLSTPLAOFREDITWRRPOICATPRLFPDDP 60
DB 1 MRAGGATPARELFRDAAPPAADSLFCDSLSTPLAOFREDITWRRPOICATPRLFPDDP 60
QY 61 REGOVKGLGDCWFLCACAALOKSRHLDOVIPPGOPSMADDEYRGSFTCRIMQFGRWY 120
DB 61 REGOVKGLGDCWFLCACAALOKSRHLDOVIPPGOPSMADDEYRGSFTCRIMQFGRWY 120
QY 121 EYTTDDRLPCLAGRLCFSCQREDVEFWLPLEKYAFAVHGSYEHLMAGVADALVDTLTCG 180
DB 121 EYTTDDRLPCLAGRLCFSCQREDVEFWLPLEKYAFAVHGSYEHLMAGVADALVDTLTCG 180
QY 181 LAERWMLKGVASGGGODRGRNEHRTCRQLHLKQCLISCCVLSFRAGARELGEHFAR 240
DB 181 LAERWMLKGVASGGGODRGRNEHRTCRQLHLKQCLISCCVLSFRAGARELGEHFAR 240
QY 241 IYSDLELQAGOCILLRIQNPWGRRCWGLMREGESMSQVDAVAASELSQLOEGE 300
DB 241 IYSDLELQAGOCILLRIQNPWGRRCWGLMREGESMSQVDAVAASELSQLOEGE 300
QY 301 FWEEBEFLREFDELTVGPVTEAGHLQSLYTERLLCHTRALPGAWKQGSAGCCRNNG 360
DB 301 FWEEBEFLREFDELTVGPVTEAGHLQSLYTERLLCHTRALPGAWKQGSAGCCRNNG 360
QY 361 FPSNPKFVLRSPESEYITAVLQSRHLHADNAGARALVGDSTMSPASITGKHYQAV 420
DB 361 FPSNPKFVLRSPESEYITAVLQSRHLHADNAGARALVGDSTMSPASITGKHYQAV 420
QY 421 GLHLMVEKRRVNLPRVLSMPVAGTACHAYDREVLKCELSGYTLAVPSTFLKDAPE 480
DB 421 GLHLMVEKRRVNLPRVLSMPVAGTACHAYDREVLKCELSGYTLAVPSTFLKDAPE 480
QY 481 FLIRVSTGRVSLRAIRAVAKTTPGALPAGEMGTVOLGSMRWGOTAGSRNFPASIPT 540
DB 481 FLIRVSTGRVSLRAIRAVAKTTPGALPAGEMGTVOLGSMRWGOTAGSRNFPASIPT 540
QY 541 NCFEPSEVGEQPRCVRITLHQHCRNPDTEPHIGFHLFOVPEGRSODAPPLLOEPL 600
DB 541 NCFEPSEVGEQPRCVRITLHQHCRNPDTEPHIGFHLFOVPEGRSODAPPLLOEPL 600
QY 601 LSCVPHRYAOEVSRLCLIPAGTYKVPSTYLPDTGCAFTVTATRIDRPSIHSEMLGOF 660
DB 601 LSCVPHRYAOEVSRLCLIPAGTYKVPSTYLPDTGCAFTVTATRIDRPSIHSEMLGOF 660
QY 661 LOEVSVMAMVKT 672
DB 661 LOEVSVMAMVKT 672
QY 506 LOEVSVMAMVKT 517
DB 506 LOEVSVMAMVKT 517

```

```

; LENGTH: 544
; TYPE: PRT
; ORGANISM: Human
; US-09-422-869-4

Query Match 74.0%; Score 2701.5; DB 4; Length 544;
Best Local Similarity 95.4%; Pred. No. 5.5e-263;
Matches 498; Conservative 5; Mismatches 6; Indels 13; Gaps 1;

QY 1 MRAGGATPARELFRDAAPPAADSLFCDSLSTPLAOFREDITWRRPOICATPRLFPDDP 60
DB 1 MRAGGATPARELFRDAAPPAADSLFCDSLSTPLAOFREDITWRRPOICATPRLFPDDP 60
QY 61 REGOVKGLGDCWFLCACAALOKSRHLDOVIPPGOPSMADDEYRGSFTCRIMQFGRWY 120
DB 61 REGOVKGLGDCWFLCACAALOKSRHLDOVIPPGOPSMADDEYRGSFTCRIMQFGRWY 120
QY 121 EYTTDDRLPCLAGRLCFSCQREDVEFWLPLEKYAFAVHGSYEHLMAGVADALVDTLTCG 180
DB 121 EYTTDDRLPCLAGRLCFSCQREDVEFWLPLEKYAFAVHGSYEHLMAGVADALVDTLTCG 180
QY 181 LAERWMLKGVASGGGODRGRNEHRTCRQLHLKQCLISCCVLSFRAGARELGEHFAR 240
DB 181 LAERWMLKGVASGGGODRGRNEHRTCRQLHLKQCLISCCVLSFRAGARELGEHFAR 240
QY 241 IYSDLELQAGOCILLRIQNPWGRRCWGLMREGESMSQVDAVAASELSQLOEGE 300
DB 241 IYSDLELQAGOCILLRIQNPWGRRCWGLMREGESMSQVDAVAASELSQLOEGE 300
QY 301 FWEEBEFLREFDELTVGPVTEAGHLQSLYTERLLCHTRALPGAWKQGSAGCCRNNG 360
DB 301 FWEEBEFLREFDELTVGPVTEAGHLQSLYTERLLCHTRALPGAWKQGSAGCCRNNG 360
QY 361 FPSNPKFVLRSPESEYITAVLQSRHLHADNAGARALVGDSTMSPASITGKHYQAV 420
DB 361 FPSNPKFVLRSPESEYITAVLQSRHLHADNAGARALVGDSTMSPASITGKHYQAV 420
QY 421 GLHLMVEKRRVNLPRVLSMPVAGTACHAYDREVLKCELSGYTLAVPSTFLKDAPE 480
DB 421 GLHLMVEKRRVNLPRVLSMPVAGTACHAYDREVLKCELSGYTLAVPSTFLKDAPE 480
QY 481 FLIRVSTGRVSLRAIRAVAKTTPGALPAGEMGTVOLGSMRWGOTAGSRNFPASIPT 542
DB 481 FLIRVSTGRVSLRAIRAVAKTTPGALPAGEMGTVOLGSMRWGOTAGSRNFPASIPT 542

RESULT 5
US-09-422-869-8
; Sequence 8, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: ODA, NAOHISA
; APPLICANT: COX, NANCY J.
; APPLICANT: SREENAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: OTANI, KENICHI
; APPLICANT: HANIS, CRAIG I.
; APPLICANT: BELL, GRAEME I.
; FILE REFERENCE: ARCD:307
; CURRENT APPLICATION NUMBER: US/09/422,869
; EARLIER FILING DATE: 1999-10-21
; EARLIER FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 513
; TYPE: PRT
; ORGANISM: Human

```

US-09-422-869-8

Query Match 74.0%; Score 2699; DB 4; Length 513;
Best Local Similarity 98.2%; Pred. No. 8.9e-263;
Matches 496; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

Qy 1 MRAGRGATPARELFDRAAFPAADSSLCFCLSTPLAQFREDITWRRPQOICATPRLFPDDP 60
Db 1 MRAGRGATPARELFDRAAFPAADSSLCFCLSTPLAQFREDITWRRPQOICATPRLFPDDP 60

Qy 61 REGOVKOGLLGDCWFLCACAALQSRHLLDQVIPPQPSWADQYRGSTCRIMQFGRW 120
Db 61 REGOVKOGLLGDCWFLCACAALQSRHLLDQVIPPQPSWADQYRGSTCRIMQFGRW 120

Qy 121 EVTTDDRLPCLAGLRCFSRCQREDVFWLPLEKYYAKVHGSYEHLMWAGQVADALVDTGG 180
Db 121 EVTTDDRLPCLAGLRCFSRCQREDVFWLPLEKYYAKVHGSYEHLMWAGQVADALVDTGG 180

Qy 181 LAERNLKVAGSGGGQDRPGRWEHRTCRQLLHLKDQCLISCCVLSPRAGARELGEHAF 240
Db 181 LAERNLKVAGSGGGQDRPGRWEHRTCRQLLHLKDQCLISCCVLSPRAGARELGEHAF 240

Qy 241 IVSDLRELQOAGOCILLRLIQRNPGRCWQGLWREGGEGHSQYDAVASELLSLOE 300
Db 241 IVSDLRELQOAGOCILLRLIQRNPGRCWQGLWREGGEGHSQYDAVASELLSLOE 300

Qy 301 FWEEEFLEFDELTGYPVTEAGHLQSLYTERLCHTRALPCAWKVGOSAGCRNNG 360
Db 301 FWEEEFLEFDELTGYPVTEAGHLQSLYTERLCHTRALPCAWKVGOSAGCRNNG 360

Qy 361 FPSNPKFWLRVSESEVYIAVLRSLHAADWAGRARALVGDSTSWSPASIPCKHYQAV 420
Db 361 FPSNPKFWLRVSESEVYIAVLRSLHAADWAGRARALVGDSTSWSPASIPCKHYQAV 420

Qy 421 GLHLWKVKRRVNLPRVLSMPPVAGTACHAYDREHVLRCELSPGYIYLA VPSFTFLKDPGE 480
Db 421 GLHLWKVKRRVNLPRVLSMPPVAGTACHAYDREHVLRCELSPGYIYLA VPSFTFLKDPGE 480

Qy 481 FLLRVFSTGRVSLSAIRAVAKNTTP 505
Db 481 FLLRVFSTGRVSLSAIRAVAKNTTP 505

RESULT 6

US-09-422-869-10
; Sequence 10, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: ODA, NAOHISA
; APPLICANT: COX, NANCY J.
; APPLICANT: SREENAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: OTANI, KENICHI
; APPLICANT: HANIS, CRAIG L.
; APPLICANT: BELL, GRAEME I.
; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
; FILE REFERENCE: ARCD:307
; CURRENT APPLICATION NUMBER: US/09/422,869
; CURRENT FILING DATE: 1999-10-21
; EARLIER APPLICATION NUMBER: 60/134,175
; EARLIER FILING DATE: 1999-05-13
; NUMBER OF SEQ. ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 10
; LENGTH: 444
; TYPE: PRT
; ORGANISM: Human
US-09-422-869-10

Query Match 64.0%; Score 2336; DB 4; Length 444;
Best Local Similarity 99.8%; Pred. No. 2.3e-226;
Matches 425; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MRAGRGATPARELFDRAAFPAADSSLCFCLSTPLAQFREDITWRRPQOICATPRLFPDDP 60
Db 1 MRAGRGATPARELFDRAAFPAADSSLCFCLSTPLAQFREDITWRRPQOICATPRLFPDDP 60

Qy 61 REGOVKOGLLGDCWFLCACAALQSRHLLDQVIPPQPSWADQYRGSTCRIMQFGRW 120
Db 61 REGOVKOGLLGDCWFLCACAALQSRHLLDQVIPPQPSWADQYRGSTCRIMQFGRW 120

Qy 121 EVTTDDRLPCLAGLRCFSRCQREDVFWLPLEKYYAKVHGSYEHLMWAGQVADALVDTGG 180
Db 121 EVTTDDRLPCLAGLRCFSRCQREDVFWLPLEKYYAKVHGSYEHLMWAGQVADALVDTGG 180

Qy 181 LAERNLKVAGSGGGQDRPGRWEHRTCRQLLHLKDQCLISCCVLSPRAGARELGEHAF 240
Db 181 LAERNLKVAGSGGGQDRPGRWEHRTCRQLLHLKDQCLISCCVLSPRAGARELGEHAF 240

Qy 241 IVSDLRELQOAGOCILLRLIQRNPGRCWQGLWREGGEGHSQYDAVASELLSLOE 300
Db 241 IVSDLRELQOAGOCILLRLIQRNPGRCWQGLWREGGEGHSQYDAVASELLSLOE 300

Qy 301 FWEEEFLEFDELTGYPVTEAGHLQSLYTERLCHTRALPCAWKVGOSAGCRNNG 360
Db 301 FWEEEFLEFDELTGYPVTEAGHLQSLYTERLCHTRALPCAWKVGOSAGCRNNG 360

Qy 361 FPSNPKFWLRVSESEVYIAVLRSLHAADWAGRARALVGDSTSWSPASIPCKHYQAV 420
Db 361 FPSNPKFWLRVSESEVYIAVLRSLHAADWAGRARALVGDSTSWSPASIPCKHYQAV 420

Qy 421 GLHLWK 426
Db 421 GLHLWK 426

RESULT 7

US-09-422-869-12
; Sequence 12, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: ODA, NAOHISA
; APPLICANT: COX, NANCY J.
; APPLICANT: SREENAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: OTANI, KENICHI
; APPLICANT: HANIS, CRAIG L.
; APPLICANT: BELL, GRAEME I.
; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
; FILE REFERENCE: ARCD:307
; CURRENT APPLICATION NUMBER: US/09/422,869
; CURRENT FILING DATE: 1999-10-21
; EARLIER APPLICATION NUMBER: 60/134,175
; EARLIER FILING DATE: 1999-05-13
; NUMBER OF SEQ. ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 12
; LENGTH: 274
; TYPE: PRT
; ORGANISM: Human
US-09-422-869-12

Query Match 23.5%; Score 858.5; DB 4; Length 274;
Best Local Similarity 81.7%; Pred. No. 4.6e-78;
Matches 165; Conservative 1; Mismatches 13; Indels 23; Gaps 3;

Qy 1 MRAGRGATPARELFDRAAFPAADSSLCFCLSTPLAQFREDITWRRPQOICATPRLFPDDP 60
Db 1 MRAGRGATPARELFDRAAFPAADSSLCFCLSTPLAQFREDITWRRPQOICATPRLFPDDP 60

0Y 61 REGVKGGLGDCWFLCACAALAKSRHLDDQVTPPGOPSMADQYRGSTFCRTIMQGRMV 120
Db 61 REGVKGGLGDCWFLCACAALAKSRHLDDQVTPPGOPSMADQYRGSTFCRTIMQGRMV 120
0Y 121 EVTTDRPLCLAGRLCFSCRCQREDVFWLPLEKRYAKVHGSYEHLMAGVADALVDTLGG 180
Db 121 EVTTDRPLCLAGRLCFSCRCQREDVFWLPLEKRYAKVHGSYEHLMAGVADALVDTLGG 180
0Y 181 LAERWNLKGYAGSGGQDRGR 202
Db 168 -----GCPGPDPRGR 179

RESULT 8
US-08-835-099A-1
Sequence 1, Application US/08835099A
Patent No. 5874277
GENERAL INFORMATION:
APPLICANT: SHINTANI, Yasushi
APPLICANT: NISHI, Kazuo
TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION
TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/835,099A
FILING DATE: 04-APR-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 083649/1996
FILING DATE: 05-APR-1996
APPLICATION NUMBER: 97105508.2
FILING DATE: 03-APR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Resnick, David S
REGISTRATION NUMBER: 34,235
REFERENCE/DOCKET NUMBER: 47342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 703 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-835-099A-1

Query Match 17.4%; Score 635; DB 2; Length 703;
Best Local Similarity 32.3%; Pred. No. 6.5e-55;
Matches 164; Conservative 85; Mismatches 179; Indels 80; Gaps 17;

0Y 13 LFRDAFPAADSSL-FCOLSTPLAQFREDITWRROEICATYRLPPDDPREGVKGGLG 71
Db 45 LFKDPEFPACPSALGKIDG-PGSPOTOGIWKRPTELCPSPOFTVGATRTDIOGGIG 103
0Y 72 DCMFLCACAALAKSRHLDDQVTPPGOPSMADQYRGSTFCRTIMQGRMVETTTDRCL 131

Db 104 DCMFLAIAISLTLLNEELLYVPPDD-FOENAGTIFHFQFMQYGMVEVVIDDLPTK 161
0Y 132 AGRICFSCRCQREDVFWLPLEKRYAKVHGSYEHLMAGVADALVDTLGGIAERWNLKGYA 191
Db 162 NGQGLFLHSGQNEFWNALLKAYAKLNGCTEALAGSTYGEFDFGTGISEFYDLK 218
0Y 192 GSGGQDPRGRWERTCROLLHLKDC--LISCV-LSPRAG--RELGFHAFIV 242
Db 219 -----KPPANLYQIIRKAL-----CAGSLGCSIDVSSAAEATITSKLVKSHAYSV 266
0Y 243 SDRELQAGAGCCILLRIQNPGRRCQGLMBEGGSGQVDAANAASELSQLQGEFFW 302
Db 267 TGVEEWNFO-GHPEKILRLNPMGEVEMSGMSDDAPEWNNHIDPRRKEELDKRVEDGEFW 325
0Y 303 VEEEFLEEFDELTVGVPVTEAGHLQSLYTERLLCHTRALPGAMVYQSGGCRNNSGFP 362
Db 326 MSLSDFYRPSRLFICNLSPDSLSEVEVHKNMLTEN---GHMTRGISTGGQN---YP 378
0Y 363 ----SNPKFWLRYSESE-----VYIAVLQSRSLHADAAGRARALVDSHTS 406
Db 379 ATWTNPOFRIRLDEVDEQEEISIGRCCVVLGLMQKN-----RRKRIGGMLS 430
0Y 407 WSPASIPGKHYQAVGLHMKVEKR---RVNLP--VLSMPVAGTACHAYDREYHLRC 459
Db 431 -----IGYAVVOVPKELESHTDAHGRDFLAYSARSFTVNLREYSGRA 477
0Y 460 ELSPGYLAVSPFELDAGEFLRYES 487
Db 478 RLPGETLVVSPTEFPKDEFLRYES 505

RESULT 9
US-09-157-349-1
Sequence 1, Application US/09157349
Patent No. 6068990
GENERAL INFORMATION:
APPLICANT: SHINTANI, Yasushi
APPLICANT: NISHI, Kazuo
TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION
TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/157,349
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/835,099.
FILING DATE:
APPLICATION NUMBER: 97105508.2
FILING DATE: 03-APR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Resnick, David S
REGISTRATION NUMBER: 34,235
REFERENCE/DOCKET NUMBER: 47342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
TELEX:
INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 703 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-157-349-1

Query Match 17.4%; Score 635; DB 3; Length 703;
Best Local Similarity 32.3%; Pred. No. 6.5e-55;
Matches 164; Conservative 85; Mismatches 179; Indels 80; Gaps 17;

QY 13 LFRDAAPRAADSS-L-FCDLSTPLAQFREDITWRPQECATRLPDDPREGVQKGLG 71
DB 45 LFKDPEFPACPSALGYKDLG-PGSPQTQGIWKRPTELCPSQFIVGGATRDICQGGIG 103
QY 72 DCWFLCAALQKSRHLLDOVIPPQPSWADQEGSTCRWQGRVVEVTTDRPLCL 131
DB 104 DCWLLAAIASLTNEELLYRVPRDQ--FOENYAGIFHFQWQYGEVVEVDDRLPTK 161
QY 132 AGRICFSRCQREDVFWPLLEKVKYAKVHSGYEHLMWAGQVADALVLTGGLAERNLKGVA 191
DB 162 NGQLFLHSEQNEFWSSALLEKAYAKLNGCYEALAGGSTVEGFTGGISEFYDLK--- 218
QY 192 GSGGQDRPGWEHTRCQLHLKDDC---LISCCV-LSPRAGA-----RELGEFHAFIV 242
DB 219 -----KPPANLYQIIRKAL-----CAGSLGCSIDVSSAAEAITSQKLVKSHAYSV 266
QY 243 SDRLEQAGQOCILLRIQNPWRCRQGLWREGGEGSWQVDAVAASELLSOLQEGEFW 302
DB 267 TGVVEVNFQ-GHPEKLIIRNPWGEVSWGAWSDDAPEWNIHIDPRRKEELDKKVEDGEFW 325
QY 303 VEREEFLREFDELTVGYPVTEAGHLQSLYTERLCHTRALPGAWVKGSAGGCRNNSGFP 362
DB 326 MSLSDVFRQPSRLEICNLSPDSLSSEEVHKNVLVFN---GHWTGRTAGGCQN---YP 378
QY 363 -----SNPKFWLRVSEPE-----VYIAVLQSRHLRAADWAGRARALYVDSHTS 406
DB 379 ATYTNFQFKIRLDEVEDQEEISGECCTVLLGLMQKNR-----RWRKRIGQGMLS 430
QY 407 WSPASIPGKHQYQVGLHLMKVEKR-----RVNLP--VLSMPPVAGTACHAYDREVLRC 459
DB 431 -----IGYAVIQVPKELESHTDAHLGRDFFLAYQPSARTSTYVNLREVSGRA 477
QY 460 ELSPGYILAVPSTFLKDPAGEFLLRVFS 487
DB 478 RLPPGEYLWVPSTFEPFKDGEFCLRVFS 505

RESULT 10
US-08-835-099A-2
Sequence 2, Application US/08835099A
Patent No. 5874277
GENERAL INFORMATION:
APPLICANT: SHINTANI, Yasushi
APPLICANT: NISHI, Kazuori
APPLICANT: KAWAMOTO, Tomohiro
TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION
TITLE OF INVENTION: AND USE
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/835,099A
FILING DATE: 04-APR-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 083649/1996
FILING DATE: 05-APR-1996
APPLICATION NUMBER: 97105508.2
FILING DATE: 03-APR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Resnick, David S
REGISTRATION NUMBER: 34,235
REFERENCE/DOCKET NUMBER: 47342
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
TELEX:

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 712 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-835-099A-2

Query Match 17.4%; Score 635; DB 2; Length 712;
Best Local Similarity 32.3%; Pred. No. 6.6e-55;
Matches 164; Conservative 85; Mismatches 179; Indels 80; Gaps 17;

QY 13 LFRDAAPRAADSS-L-FCDLSTPLAQFREDITWRPQECATRLPDDPREGVQKGLG 71
DB 54 LFKDPEFPACPSALGYKDLG-PGSPQTQGIWKRPTELCPSQFIVGGATRDICQGGIG 112
QY 72 DCWFLCAALQKSRHLLDOVIPPQPSWADQEGSTCRWQGRVVEVTTDRPLCL 131
DB 113 DCWLLAAIASLTNEELLYRVPRDQ--FOENYAGIFHFQWQYGEVVEVDDRLPTK 170
QY 132 AGRICFSRCQREDVFWPLLEKVKYAKVHSGYEHLMWAGQVADALVLTGGLAERNLKGVA 191
DB 171 NGQLFLHSEQNEFWSSALLEKAYAKLNGCYEALAGGSTVEGFTGGISEFYDLK--- 227
QY 192 GSGGQDRPGWEHTRCQLHLKDDC---LISCCV-LSPRAGA-----RELGEFHAFIV 242
DB 228 -----KPPANLYQIIRKAL-----CAGSLGCSIDVSSAAEAITSQKLVKSHAYSV 275
QY 243 SDRLEQAGQOCILLRIQNPWRCRQGLWREGGEGSWQVDAVAASELLSOLQEGEFW 302
DB 276 TGVVEVNFQ-GHPEKLIIRNPWGEVSWGAWSDDAPEWNIHIDPRRKEELDKKVEDGEFW 334
QY 303 VEREEFLREFDELTVGYPVTEAGHLQSLYTERLCHTRALPGAWVKGSAGGCRNNSGFP 362
DB 335 MSLSDVFRQPSRLEICNLSPDSLSSEEVHKNVLVFN---GHWTGRTAGGCQN---YP 387
QY 363 -----SNPKFWLRVSEPE-----VYIAVLQSRHLRAADWAGRARALYVDSHTS 406
DB 388 ATYTNFQFKIRLDEVEDQEEISGECCTVLLGLMQKNR-----RWRKRIGQGMLS 439
QY 407 WSPASIPGKHQYQVGLHLMKVEKR-----RVNLP--VLSMPPVAGTACHAYDREVLRC 459
DB 440 -----IGYAVIQVPKELESHTDAHLGRDFFLAYQPSARTSTYVNLREVSGRA 486
QY 460 ELSPGYILAVPSTFLKDPAGEFLLRVFS 487
DB 487 RLPPGEYLWVPSTFEPFKDGEFCLRVFS 514

RESULT 11
US-09-157-349-2
Sequence 2, Application US/09157349
Patent No. 6068990
GENERAL INFORMATION:

APPLICANT: SHINTANI, Yasushi
 APPLICANT: NISHI, Kazuo
 APPLICANT: KAWAMOTO, Tomohito
 TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION
 TITLE OF INVENTION: AND USE
 NUMBER OF SEQUENCES: 18
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
 STREET: 130 Water Street
 CITY: Boston
 STATE: MA
 COUNTRY: USA
 ZIP: 02109
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/157,349
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/835,099
 FILING DATE:
 APPLICATION NUMBER: 97105508.2
 FILING DATE: 03-APR-1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Resnick, David S.
 REGISTRATION NUMBER: 34,235
 REFERENCE/DOCKET NUMBER: 47342
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617-523-3400
 TELEFAX: 617-523-6440
 TELEX:
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 712 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: Protein
 US-09-157-349-2

Query Match 17.4%; Score 635; DB 3; Length 712;
 Best Local Similarity 32.3%; Pred. No. 6.6e-55;
 Matches 164; Conservative 85; Mismatches 179; Indels 80; Gaps 17;

13 LFRDAFPAADSSL-FCDDSTPLAQRREITRRPOEICATPRLFPDDPREGQVKGGLG 71
 54 LKRDPEFPSPALGKIDG-PCSPOTGIIWKRPTELCPSPQFIYGATRTDIOGGG 112
 72 DCMFLCACAALOKSRHLDQVTPPGPSWADQYRSGFTCRIMQGRWVETTDRLPCL 131
 113 DCMFLAALASLFLNELLYLRYVPRDQ-FOENYAGIFHFQWQGEVYIDRLPFRK 170
 132 AGRCLFRCQRQEDVFWPLLEKYAVKHGSEYELHAGVADALVDTGLGLAEKMNKGYA 191
 171 NGDLFLHSEQNEFMSALKEKAYAKLNGCYALAGSTVEGFEDFTGISEFYDK--- 227
 192 GSGGQDDPRGRHRTCRQLHLKQDC---LISCV-LSPRAGA---RELGEFHAFLV 242
 228 -----KPMNLVQIIRKAL-----CAGSLGCSIDVSAEAERISQKLYKSHASV 275
 243 SFLRELQAGOCILLRLIOWPGRRCWGLMREGGEGMSQVDAAVASSELLQLOEGEFW 302
 276 TGEVAVNQ-GHEPKLRLKRWPGEVMSGAMSDAPMNMHIDPRRKEELDKVDEGEFW 334
 303 VEEEFLEFDELIVGYVTEAGHLQSLYTERLCHTRALPGAWKQSGAGCCRNNSGFP 362
 335 MSLSDVFWQFRLTEICNLSPSLSSEVYHAKMNLVFN-----GHWTGRSTAGCCN---YP 387

363 -----SNPKFWLVRSESE-----VYIAVLQSRLLHAAADWAGARALVGSHTS 406
 388 ATYTNPOKRIKRLDEDEQESIGPCTVLLGLMKR-----HMRKRIQGMIS 439
 407 WSPASIPGRHYQVGLHKKYER-----RVNLP--VLSMPVAGTACHADREYHLK 459
 440 -----IGAYIVPKELSEHTDAHGRFLAYOPASATSTYVNLREVSGRA 486
 460 ELSPEYIAVPSTFLKDAPEFLRYFS 487
 487 RLPPGEYLVPTSPFPKDEGCELVFS 514

RESULT 12
 US-09-422-869-22
 Sequence 22, Application US/09422869
 Patent No. 6235481
 GENERAL INFORMATION:
 APPLICANT: POLONSKY, KENNETH S.
 APPLICANT: HORIKAWA, YUKIO
 APPLICANT: ODA, NAOHISA
 APPLICANT: COX, NANCY J.
 APPLICANT: SREENAN, SEAMUS
 APPLICANT: ZHOU, YUN-PING
 APPLICANT: OTANI, KENICHI
 APPLICANT: HANIS, CRAIG L.
 APPLICANT: BELL, GRAEME I.
 TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
 FILE REFERENCE: ARCD:307
 CURRENT APPLICATION NUMBER: US/09/422,869
 EARLIER FILING DATE: 1999-10-21
 EARLIER FILING DATE: 1999-05-13
 NUMBER OF SEQ ID NOS: 30
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 22
 LENGTH: 714
 TYPE: PRT
 ORGANISM: Human
 US-09-422-869-22

Query Match 16.9%; Score 618.5; DB 4; Length 714;
 Best Local Similarity 32.3%; Pred. No. 3e-53;
 Matches 166; Conservative 73; Mismatches 198; Indels 77; Gaps 17;

13 LFRDAFPAADSSL-FCDDSTPLAQRREITRRPOEICATPRLFPDDPREGQVKGGLG 71
 55 LFRDAFPPVPSGLGKIDG-PNSSKTYGKIKRPELISNPOFIYGATRTDIOGALG 113
 72 DCMFLCACAALOKSRHLDQVTPPGPSWADQYRSGFTCRIMQGRWVETTDRLPCL 131
 114 DCMFLAALASLFLNELLYLRYVPHGOS--FQNGYAGIFHFQWQGEVYIDRLPFRK 171
 132 AGRCLFRCQRQEDVFWPLLEKYAVKHGSEYELHAGVADALVDTGLGLAEKMNKGYA 191
 172 DKLIVFVHSAEENEFMSALKEKAYAKLNGCYALAGSTVEGFEDFTGISEFYDK--- 221
 192 GSGGQDDPRGRHRTCRQLHLKQDC---CYLSRA-GARELGEFHAFLVSD 245
 232 SLYQ-----ILKALDERGSLGCSIDISSVLDMAEITFRKLVKGAHSVTA 279
 246 RELQAGOCILLRLIOWPGRRCWGLMREGGEGMSQVDAAVASSELLQLOEGEFW 305
 280 KQV-NYRGQVSLIRBRNPGEVMSGAMSDSSSENNVDPERDQLRYKMEDEGEFWSF 338
 306 EEFLEFDELIVGYVTEAGHLQSLYTERLCHTRALPGAWKQSGAGCCRNNSGFP 363
 339 RDMFRELFRLEICNLTLPDA--LKSRTIRK--WNTLLYEGTWRGRTAGCCRN---YPATF 391
 364 --NPKFWLVRSESE-----VYIAVLQSRLLHAAADWAGARALVGSHTS 409
 392 WVNPOKRIKRLDEDDYDGRSGCSFVLAALMKR-----RRERF----- 434

QY 410 ASIPGKHQVAGLHMLKVEKRRVNLPRV-LSMPPVAGTACHAYD-----REVHLRCELS 462
Db 435 ---GRDMETIGVAYEVEPELVGQPAVHLKRDFFLANASRASEQFINLREVSTRFLP 490
QY 463 PGYILAVPSTFKDAPGEFLRVF---STGRVSL 493
Db 491 PGEYVVVSTFPNKEGDFVLRFFSEKSAGTVEL 524

RESULT 13
US-09-422-869-24
: Sequence 24, Application US/09422869
: Patent No. 6235481
: GENERAL INFORMATION:
: APPLICANT: POLONSKY, KENNETH S.
: APPLICANT: HORIKAWA, YUKIO
: APPLICANT: ODA, NAOHISA
: APPLICANT: COX, NANCY J.
: APPLICANT: SREENAN, SEAMUS
: APPLICANT: ZHOU, YUN-PING
: APPLICANT: OTANI, KENICHI
: APPLICANT: HANIS, CRAIG L.
: APPLICANT: BELL, GRAEME I.
: TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
: FILE REFERENCE: ARCD:307
: CURRENT APPLICATION NUMBER: US/09/422.869
: CURRENT FILING DATE: 1999-10-21
: EARLIER APPLICATION NUMBER: 60/134.175
: EARLIER FILING DATE: 1999-05-13
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 24
: LENGTH: 821
: TYPE: PRT
: ORGANISM: Human
US-09-422-869-24

Query Match 16.9%; Score 618; DB 4; Length 821;
Best Local Similarity 31.4%; Pred. No. 4.3e-53;
Matches 178; Conservative 79; Mismatches 194; Indels 116; Gaps 19;
QY 13 LFRDAAFPADSSIFCDLSTPLAQFREDITWRPQOEICATPRLPDPPREGQVQGLGD 72
Db 74 LYVPEFPDPTSLFYQKFPPI-QF---VWKRPEICEENPRFIDGANTRDICQGLGD 128
QY 73 CWFCLACAAALQKSRHLIDQVPPGQPSWADQYRGSETCRIWQGRVWVTTDDRLPCL 132
Db 129 CWFCLAACTLNLHLLFRVIPHQS--FIENYAGIFHFQWRYGEWVVDVDDCLPTYN 186
QY 133 GRICFSRCQREDVFWLPLEKVIYAKVHGSYEHLMWAGQVADALVDLTGGLA----- 182
Db 187 NOLVETKSNHRNEFWFSALEKAYAKLHGSYEALKGNTTEAMEDFTGVAEFFEIRDAPS 246
QY 183 -----ERNWLGVA-----GSGQGDPRGRWEHRTCRQLLHLKDOCLISCC 223
Db 247 DMYKIMKATIERGSLMGCSIDGNTMTYGTSPGLNMGELIARVRNN-----DNSLLQDS 302
QY 224 VLSPRAG-----ARELGEFHAFIVSDRLQOGAOCITLLRIQNPWG 266
Db 303 DLDPRGSDERTRTIIPVQYETRMACGLVRCHAYSVTGLDEVPK-GEKVLVRLRNPWG 361
QY 267 RRCQGLWREGGEGSWQDAVASELLSQL-QEGEFWEEEFLEFDELTVGYPVTEAG 325
Db 362 QVEWNGSWSRDKWDFVDEKARLQHVTEDEGEFWSYEDFIYHFTKLEICNLTA- 420
QY 326 HQSLYTERLICHTRAL-PCGAWKVGQSGGCRNNSGFP-----SNPKFWLRVSE-----P 374
Db 421 ----LQSDKLTQTTVSVNEGRWRCGSGAGCRN---FPDFTWNTNPOYRLKLLLEDDDDPD 473
QY 375 SEV-----YIAVLQSRSLHAADWAGARALVGDSDHTSNPASIPGKHQVAGLHMLWKV- 427

Db 474 SEVICSFLVALMQNR-----RKDRKL-----GASLFTIGFAIYEVPE 512
QY 428 ---EKRRVNLPRVLSMPPVAGTACHAYDREVHLRCELSPGYILAVPSTFKDAPGEFLR 484
Db 513 MHGNKQHLQKDFLNLASKARSKTYINMREVSQRFRLPPSEYVIVPSTYEPHQEGEFLR 572
QY 485 VFSTGR-----VLSAIRAVAKNTT 504
Db 573 VFSEKRNLSSEVENTISVDRPVKKKKT 599

RESULT 14
US-09-422-869-27
: Sequence 27, Application US/09422869
: Patent No. 6235481
: GENERAL INFORMATION:
: APPLICANT: POLONSKY, KENNETH S.
: APPLICANT: HORIKAWA, YUKIO
: APPLICANT: ODA, NAOHISA
: APPLICANT: COX, NANCY J.
: APPLICANT: SREENAN, SEAMUS
: APPLICANT: ZHOU, YUN-PING
: APPLICANT: OTANI, KENICHI
: APPLICANT: HANIS, CRAIG L.
: APPLICANT: BELL, GRAEME I.
: TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
: FILE REFERENCE: ARCD:307
: CURRENT APPLICATION NUMBER: US/09/422.869
: CURRENT FILING DATE: 1999-10-21
: EARLIER APPLICATION NUMBER: 60/134.175
: EARLIER FILING DATE: 1999-05-13
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 27
: LENGTH: 703
: TYPE: PRT
: ORGANISM: RAT
US-09-422-869-27

Query Match 16.8%; Score 612; DB 4; Length 703;
Best Local Similarity 31.2%; Pred. No. 1.3e-52;
Matches 155; Conservative 87; Mismatches 197; Indels 58; Gaps 14;
QY 13 LFRDAAFPADSSIFCDLSTPLAQFREDITWRPQOEICATPRLPDPPREGQVQGLLG 71
Db 45 LFRDPEFPAPCSALGYKDLG-PGSPDQGVIVMKRPTLCPNPFQIVGGATFDRQGLG 103
QY 72 DCWFLCACAALQKSRHLIDQVPPGQPSWADQYRGSETCRIWQGRVWVTTDDRLPCL 131
Db 104 DCWLLAAIASITLNEKLYRVLPRDQS--FKDYAGIFHFQWRYGEWVVIDRLPTK 161
QY 132 AGRICFSRCQREDVFWLPLEKVIYAKVHGSYEHLMWAGQVADALVDLTGGLAERNLKGVA 191
Db 162 NGQLLFHSEEGNEFWFSALEKAYAKLHGSYEALVGGSTIEGFEFTGGISEFYDLK--- 218
QY 192 GSGQGDPRGRWEHRTCRQLLHLKDOCLISCCV-LSPRAG-----RELGEFHAFIVSD 245
Db 219 -----KPENLYYIYIQA--LKGSLGCSIDVSTAAEATROKLVAGHAYSVTV 269
QY 246 RELQOGAOCITLLRIQNPWRCWQGLWREGGEGSWQDAVASELLSQLQEGEFWEE 305
Db 270 EEVNFH-GRPEKLIRLNPWGEVWSCAWSDNAPEWNYIDPRKEELDRAEDGEFWSF 328
QY 306 EELREFDELTVGYPVTEAGHLSLYTERLICHTRALPGAWKVGQSGAGGCRNNSG-FPSN 364
Db 329 SDFLKQYSRLEICNLSPDSLSESEIHKWNLVFN---GRWTRGTAGGCLNPGTYWTN 384
QY 365 PKFWLRVSEPE-----YIAVLQSRSLHAADWAGARALVGDSDHTS--WSPA 410
Db 385 PQKIHLDVEDEQEEGTSEPCCTVLLGLMQNR-----RROKRIGOGMLSIGTAVY 436
QY 411 SIPGKHQVAGLHMLKVEKRRVNLPRVLSMPPVAGTACHAYDREVHLRCELSPGYILAVP 470

Db 437 QIPKELESHTDAHLGR-----DEFLLGRPSTCSSTVMNLRVSSRRRLPGQLYVP 488
 QY 471 STEFLDAPGEFLLRVES 487
 Db 489 STEEPKDDPCLLRVES 505

RESULT 15

US-08-726-525-7

Sequence 7, Application US/08726525

Patent No. 5789181

GENERAL INFORMATION:

APPLICANT: Lin, Lin-Ling

APPLICANT: Graham, James

TITLE OF INVENTION: NOVEL INTERLEUKIN-1 RECEPTOR

TITLE OF INVENTION: INTRACELLULAR LIGAND PROTEINS AND INHIBITORS OF LIGAND

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESSES:

ADDRESSEE: LEGAL AFFAIRS, GENETICS INSTITUTE, INC.

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: MA

COUNTRY: USA

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/726,525

FILING DATE: 07-OCT-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/487,942

FILING DATE: 07-JUN-1995

ATTORNEY/AGENT INFORMATION:

NAME: Brown, Scott A.

REGISTRATION NUMBER: 32,724

REFERENCE/DOCKET NUMBER: G15258

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 498-8224

TELEFAX: (617) 876-5851

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 700 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

US-08-726-525-7

Query Match

Best Local Similarity 16.6%; Score 604; DB 1; Length 700;

Matches 157; Conservative 76; Mismatches 209; Indels 74; Gaps 15;

QY 10 ARELFRDAPFADSSL-FCDDLSTPLAQFREDITWRROICATPRLFPPDPREGOVKOG 68
 Db 42 AGTLFDPSPPALPSALGFELG-PTSSKTRGKMKRPTICADPOFIIGATRTDICO 100
 QY 69 LLDCCWFLCACAIAKSRHLLDVIPGGQSMADQETRGSTFCRIMQFGRWVEVTTDDRL 128
 Db 101 ALGDCWFLAIAIASLTLNEELAVVPLNOS--FOENYAGIFHFQFMOYGEVVEYVDDRL 158
 QY 129 PCLAGRLCFSRQREDVFNPLLEKYAKYVHSYEHLMAGOVADALVLTGGLAERWNLK 188
 Db 159 PTYDGLLEFVHSAEGSEFSALEKAYAKINGCYEALSGATGEGFEDFTGIAEWYELK 218
 QY 189 GVAGSGGQDPRGKWHRCRQLHLKDCLLSCVLSPRAGARELGEF-----HAFTV 242

Db 219 -----KPPNLFKTIQKA--LQKSLGCSIDITSDADSSEAITFOKLYKGHAYSV 266
 QY 243 SDLRELQAGOCICILLRIQNPGRRCWQGLMREGESQVDAVAABELLSQLOEGEW 302
 Db 267 TGAEEVEEN-GLQKLRIRNPMGEVETGRNNDCCPSNITIDPEPERLRRHEDGEW 325
 QY 303 VEEEFLEFDELTVGVPTVTEAGHLQSLYTERLCHT-----RALPGAWYQSGAGGR 356
 Db 326 MSFSDFLRHSRL-----EICNLEPDTLFTSDTYKKMKLTKMDGNWRRGSTAGGR 375
 QY 357 N-NSGFSPNPKFWLRVSEPE-----YIIVLQSRSLHAADWAGRARALVDSHT 405
 Db 376 NYPNTEWNNPOLYLRKLEEDDEDEDEDESGCTFLVGLQKHR-----RROKMGEDMHT 428
 QY 406 -SMSPASIPGKHVQAVGLMK-----VEKRRVNLPRVLSMPVAGTACHADREVLHCE 460
 Db 429 IGFGIYVPEBELSGQTNHLSKNFPLTRARERSPTFTNL-----REVLRPK 476
 QY 461 LSPGIYLAVPSTFLKADAPGEFLLRVSTGRVSLSAI 496
 Db 477 LPPEGYILVPTSPFEPNKDGCIRVSEKADYQAV 512

Search completed: September 19, 2002, 14:49:17
 Job time: 26 sec

us-09-768-877-2.sepl9.rag

20

GenCore version 4.5
(c) 1993 - 2000 Com

2152.295 Million cell updates/sec

...SØEMIGØFLØEVSVMAVMT 672

747574

Data/genseq_emb1/A1980.DAT
 Data/genseq_emb1/A1981.DAT
 Data/genseq_emb1/A1982.DAT
 Data/genseq_emb1/A1983.DAT
 Data/genseq_emb1/A1984.DAT
 Data/genseq_emb1/A1985.DAT
 Data/genseq_emb1/A1986.DAT
 Data/genseq_emb1/A1987.DAT
 Data/genseq_emb1/A1988.DAT
 Data/genseq_emb1/A1989.DAT
 Data/genseq_emb1/A1990.DAT
 Data/genseq_emb1/A1991.DAT
 Data/genseq_emb1/A1992.DAT
 Data/genseq_emb1/A1993.DAT
 Data/genseq_emb1/A1994.DAT
 Data/genseq_emb1/A1995.DAT
 Data/genseq_emb1/A1996.DAT
 Data/genseq_emb1/A1997.DAT
 Data/genseq_emb1/A1998.DAT
 Data/genseq_emb1/A1999.DAT
 Data/genseq_emb1/A2000.DAT
 Data/genseq_emb1/A2001.DAT

ected by chance to have a
of the result being printed
score distribution.

[illegible]

Novel	human protei	Novel	human protei
Human	ORFX ORF65	Human	ORFX ORF65
Novel	human protein SEQ	Novel	human protein SEQ
Novel	human protein	Novel	human protein
Novel	testis-specific	Novel	human prote
Novel	human prote	Novel	human prote
Human	calpain, h	Human	calpain, h
Human	ORFX ORF154	Human	ORFX ORF154
Human	calpain, H	Human	calpain, H
Human	calpain, R	Human	calpain, R
Calpain	large sub	Calpain	large sub
Mouse	truncated c	Mouse	truncated c
Mouse	calpain nC	Mouse	calpain nC
Mouse	calpain nC	Mouse	calpain nC
Mouse	calpain nC	Mouse	calpain nC
Rat	calpain 80kDa	Rat	calpain 80kDa
Human	calpain 80k	Human	calpain 80k
A calpain	protein	A calpain	protein
Human	ORFX ORF952	Human	ORFX ORF952
Human	CANF used f	Human	CANF used f
Human	interleukin	Human	interleukin
Human	calpain nC	Human	calpain nC
Drosophila	melanoc	Drosophila	melanoc
Drosophila	melanoc	Drosophila	melanoc
Human	calpain 10g	Human	calpain 10g
Novel	human disgn	Novel	human disgn
Novel	human enzym	Novel	human enzym
Human	polypeptide	Human	polypeptide
Mouse	calpain CAP	Mouse	calpain CAP
Human	calpain 10h	Human	calpain 10h
Drosophila	melanoc	Drosophila	melanoc

0 gene;
diagnosis;
Otani K;
rises detecting a

PT polymorphism in a calpain encoding nucleic acid segment or a
PT protease-encoding nucleic acid segment -
XX Claim 40; Page 217-219; 257pp; English.
PS The present sequence is that of a novel human cysteine protease,
XX designated calpain 10a. The protein is encoded by the major
transcript (see AA27476) of the human calpain 10 gene, CAPN10 (see
AA27475), which has been identified in the NIDDM1 region of
chromosome 2. The CAPN10 gene consists of 15 exons spanning 32 kb.
A complex pattern of alternative splicing generates proteins of
672, 544, 517, 513, 444, 274, 139 and 138 amino acids (see
AA79567-74), designated calpain 10a to 10h, respectively. The
invention concerns the identification of genes responsible for
type 2 diabetes for use in diagnostic and therapeutic applications.
A G-to-A polymorphism in intron 3 of the CAPN10 gene (UCSNP-43)
shows evidence for linkage to type 2 diabetes. Claimed methods for
screening for a propensity for type 2 diabetes mellitus are based
on detection of a polymorphism in a calpain encoding nucleic acid,
especially UCSNP-10 of the CAPN10 gene. Methods are also claimed
for: producing calpain 10 polypeptides using calpain 10-encoding
polynucleotides; regulating diabetes by providing a calpain
polypeptide such as calpain 10a; identifying modulators of calpain
activity using calpain 10 polypeptides; and using these modulators
to treat diabetes, in particular through the regulation of an
insulin secretory response or insulin mediated glucose transport.
XX Sequence 672 AA;

CC	Query Match	100.0%;	Score 3649;	DB 21;	Length 672;
CC	Best Local Similarity	100.0%;	Pred. No. 0;		
CC	Matches 672;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
CC	1 MRACRGATPARELFDAFAADSSLCFCDLSTPLAOPREDITWRROEICATRLFPDDP	60			
CC	1 mragrgatparelfrdaafpaadsslfcldstplagfreditwrrpqlcatrlfpddp	60			
CC	61 REGGVKGLGDCWFLCACAALAKSRLLDQVPPGQPSWADQEIYRGSFTCRWQFGRW	120			
CC	61 reggvkgllgdcwflcacaalqksrllhldqvppgqpswadqeyrgsftrwqfgrw	120			
CC	121 EVTTDDRLPLCLACFSCQREDFVFLPLEKVKYAKVHGSYEHVLWAGQVADALVDLTGG	180			
CC	121 evttddrlplclacfsrqredvflplekvyakvghsyehvlwagvqvadaldvltgg	180			
CC	181 LAERNLKVAGSGGQDDRPGRWEHRTCTROLLKDKQCLISCCVLSFRAGARELGEFHF	240			
CC	181 laernlkgvaggsggqddrpgrrwehrtcrqlhldkqclisccvlsfragarelgfghf	240			
CC	241 IVSDRLQOAGCCILLRIQNPWRRCWQGLWRGEGGWSQVDAVAASELLSQLOEGE	300			
CC	241 ivsdrlqgagccillirigpwrircwqglwreggegwsqvdaavasellslqleqe	300			
CC	301 FWVEEEFLREFDELTVGYPTVTEAGHLSLYTERLLCHTRALPGAVKQSGAGCCRNNSC	360			
CC	301 fwveeeeflrefdeltvgyptvteaghlsllyterllchtralpgavkqsgagccrnns	360			
CC	361 FPNPKFWLRVSEPEYVIYAVLQSRHLHADWAGARALVGDSTWSPASIPKHYQAV	420			
CC	361 fspnkwfwrsvsepeyviylavlsrhlhadwagraralvgdstwspsaspikhyqav	420			
CC	421 GLHLWKEKRVNLPRVLSPPVAGTACHAYDREHRLCELSPGYLAVPSTFLKDAPGE	480			
CC	421 glhlwkekrvnlprvlsppvagtachaydrewhrlcelspgyylavpstflkdapge	480			
CC	481 FLIRVFTGSRVLSAIRAVAKNTTPGAALPAGSGWTVQLRGSWRVGTAGGSNFASTPT	540			
CC	481 flirvftgsvlsairavaknttpgaalpagsgwvqlrgswrvgtaggsnfnstpt	540			
CC	541 NPCFPSPVPGPCRVITLHQHCRPSDTEHPHIFQVPEGGRSQDAPPLLLQEP	600			
CC	541 npcftspvpgpcrvitlhbqcrpsdtefphifqvpegrsqdappllllepl	600			

OY	601	LSCVPHRYAOEVSRLCLLPAGTYKVPSTYLPDTGEGFTVTIATRIDRPSHSGMLGOF	660
Db	601	lscvphryaqevsrlcllpagtykvpstypdtcegftvtiatridrpslshsgmlgof	660
OY	661	LQEVSVMAVMKT	672
Db	661	lqevsvmavmkt	672
RESULT	2		
AAAY70019			
ID	AAAY70019	standard; Protein; 672 AA.	
XX	AAAY70019;		
XX	05-JUN-2000	(first entry)	
XX	Human Protease and associated protein-13 (PPRG-13).		
XX	Protease and associated protein-13; PPRG-13; anti-PPRG antibody;		
XX	diagnosis; treatment; cell proliferative disorder; cancer; cirrhosis;		
XX	arteriosclerosis; atherosclerosis; bursitis; hepatitis; immune disorder;		
XX	AIDS; Addison's disease; adult respiratory distress syndrome; allergy;		
XX	ankylosing spondylitis; amyloidosis; cytostatic; antiarteriosclerotic;		
XX	hepatotrophic; antiinflammatory; virucide; antipsoriatic; anti-HIV;		
XX	antiallergic; immunosuppressive; antidiabetic; antianemic;		
XX	neuroprotective; human.		
OS	Homo sapiens.		
XX	Key	Location/Qualifiers	
FT	Region	13..322	
FT	Region	/note= "Signature sequence of Calpain"	
FT	Region	57..78	
FT	Region	/note= "Signature sequence of Cysteine protease"	
FT	Modified-site	99	
FT	Modified-site	/note= "Potential phosphorylation site"	
FT	Modified-site	123	
FT	Modified-site	/note= "Potential phosphorylation site"	
FT	Modified-site	282	
FT	Modified-site	/note= "Potential phosphorylation site"	
FT	Modified-site	547	
FT	Modified-site	/note= "Potential phosphorylation site"	
FT	Modified-site	568	
FT	Modified-site	/note= "Potential phosphorylation site"	
FT	Modified-site	644	
FT	Modified-site	/note= "Potential phosphorylation site"	
FT	Modified-site	42	
FT	Modified-site	/note= "Potential phosphorylation site"	
FT	Modified-site	52	
FT	Modified-site	/note= "Potential phosphorylation site"	
FT	Modified-site	110	
FT	Modified-site	/note= "Potential phosphorylation site"	
FT	Modified-site	207	
FT	Modified-site	/note= "Potential phosphorylation site"	
FT	Modified-site	226	
FT	Modified-site	/note= "Potential phosphorylation site"	
FT	Modified-site	332	
FT	Modified-site	/note= "Potential phosphorylation site"	
FT	Modified-site	488	
FT	Modified-site	/note= "Potential phosphorylation site"	
FT	Modified-site	522	
FT	Modified-site	/note= "Potential phosphorylation site"	
FT	Modified-site	622	
FT	Modified-site	/note= "Potential phosphorylation site"	
XX	WO200009709-A2.		
XX	24-FEB-2000.		
XX	06-AUG-1999;	99WO-US17818.	

PR 10-AUG-1998; 98US-0096114.
PR 11-FEB-1999; 99US-0119768.
XX
XX
XX (INCYTE PHARM INC.)

PI Bandman O, Hillman JT, Baughn MR, Azimzai Y, Guegler KT;
PI Corley NC, Yue H, Tang YT, Reddy R, Patterson C, Au-Young J;
PI Shih LT, Lu DAM;

XX WPI: 2000-224346/19.
XX N-PSDB: AAK50930.

PT New human proteases, useful for diagnosis, treatment and prevention of
PT cell proliferative disorders such as atherosclerosis -
XX
XX Claim 1; Page 89-90; 114pp: English.

CC The present sequence is that of human protease and associated protein-13
CC (PPRG-13), which is expressed in nervous and reproductive tissues. It
CC is encoded by cDNA identified in Incyte clone 1902576 derived
CC from OVARNO07 cDNA library. Anti-PPRG antibodies can be used
CC as therapeutic antagonists, reagents for diagnosis and monitoring
CC diseases and for isolating PPRG. PPRG nucleotide sequence can be used
CC as probe or primer for diagnosis and monitoring of PPRG-related
CC diseases and gene mapping. PPRG can be used in the treatment of cell
CC proliferative disorders like cancer, arteriosclerosis, atherosclerosis,
CC bursts, cirrhosis and hepatitis, and immune disorders like AIDS,
CC Addison's disease, adult respiratory distress syndrome, allergies,
CC ankylosing spondylitis and amyloidosis.

XX Sequence 672 AA;

Query Match 99.9%; Score 3644; DB 21; Length 672;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 671; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 MRAGRGATPARELFRDAAPADSSLCFCDLSTPLAOFREDITWRROEICTATPRLPDDP 60
DB 1 mragrgatparelfrdaapadsslfcfcdlstplagfireditwrroelcatprlfpddp 60
OY 61 REGGVKGLGDCWFLCACAALAKRSRLDQVIRPGOPSNADQYRSCFCRIMQFGRWV 120
DB 61 reggvkglgdcwflcacaalqkrsrlhldqvirpgpsnadqyrsfcrimqfgrwv 120
OY 121 EVTTDDRLPCLAGRLCFSCQREDVFWLPLEKVAHVHSGYELHMGVADALVDTLGG 180
DB 121 evttddrlpclagrlcfscqredvfwlplekvaahvshgylhmgvadaldvdtlvg 180
OY 181 LAERWNLKGVAGSGGODRGRHWRHRCROLHLKQCLISCCVLSRAARAEIHFHAF 240
DB 181 laerwnlkgvagsggodrgrhwrhrcrolhlkqclisccvlsraaraeihfhaf 240
OY 241 IVSDRLRLQAGQACILILRIQNPWGRRCWGLMREGGEGMSQVADAVASEILSQLQGE 300
DB 241 ivsdrlrlqagqacililriqnpwgrrcwglmreggegmsqvdavaseilsqlqge 300
OY 301 FWVEEERFLERDELTVGVVTEAGHLQSLYTERLLCHTRALPGAWKGGSGCCRNNSG 360
DB 301 fwveeerflerdeltvgyvteaghlqsllyterllchtralpawkgsgsgccrnns 360
OY 361 FPNPKFRLRVSESEYIVLVQSRRLHADNMGARALVGDSTMSRSTGKHQAV 420
DB 361 fpnpkfrlrvseseyivlvqsrhlhadnmgaralvgsdmsrstgkhqav 420
OY 421 GLHLAKREKRRVNLPRVLSNPVAGTACHAYDREVLKCEISPGYTLAVSTFLKADPGE 480
DB 421 glhlakrekrrvnlprvlsnppvagtachaydrevlkcelispgytlavstflkadge 480
OY 481 FLLRVSTGVSAIRAVAKNTTPGALPAGMGTVQLGSMRVGOTAGGSNRFASPT 540
DB 481 flrvstgvsaairavaknttpgalpagmgtvqlgsmrvvgotaggsnrfaspt 540

OY 541 NPCEFSVPEGPGPCVAVITLHOCRPSDTEHPHIGHIFOVPEGCSODAPPELLQEP 600
DB 541 npcfsvpepgpgpcvavitlhocrpsdtehpghihifovpegcsodappellqep 600
OY 601 LSCVPHRYAOFESRLCLPAGYKRVVPTSTYLPDEGACTVLTARPIDRPSHSQEMLCOF 660
DB 601 lscvphryaofesrlclpagykrvvptstylpdegactvltarpidrpsshsqemlcof 660
OY 661 LQEVSVMAVMKT 672
DB 661 lqevsvmavmkt 672

RESULT 3
AAM93341
ID AAM93341 standard; Protein: 672 AA.

AC AAM93341;
XX
XX 06-NOV-2001 (first entry)

DE Human polypeptide, SEQ ID NO: 2881.

KM Human: full length cDNA; cDNA synthesis; oligo-capping.

OS Homo sapiens.

PN EP1130094-A2.

PD 05-SEP-2001.

PE 07-JUL-2000; 2000EP-0114089.

PR 08-JUL-1999; 99JP-0194486.

PR 11-JAN-2000; 2000JP-0118774.

PR 02-MAY-2000; 2000JP-0183765.

XX (HELI-) HELIX RES INST.

PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Makamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

DR WPI: 2001-524255/58.

DR N-PSDB: AAK94261.

PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -

XX Claim 8; SEQ ID NO 2881; 1380bp + sequence listing; English.

XX The invention relates to primers for synthesizing full length cDNA
XX clones. 830 cDNA molecules encoding a human protein have been
XX isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
XX molecules have been determined. Primers for synthesizing the full length
XX cDNA are useful for clarifying the function of the protein encoded by
XX the cDNA. The full length clones were obtained by construction of full
XX length enriched cDNA libraries that were synthesised by the oligo-capping
XX method. The primers enable the production of the full length cDNA easily
XX without any special methods. The present sequence is a polypeptide
XX encoded by a full length human cDNA of the invention.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in CD-ROM format directly from EPO.

XX Sequence 672 AA;

Query Match 99.7%; Score 3638; DB 22; Length 672;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 670; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 MRAGRGATPARELFRDAAPADSSLCFCDLSTPLAOFREDITWRROEICTATPRLPDDP 60
DB 1 mragrgatparelfrdaapadsslfcfcdlstplagfireditwrroelcatprlfpddp 60

QY 61 REGOVKGLLGDWFLCACAALOKSRHLLDVIPPGQPSWADQYRGSTCRWQFGRWV 120
Db 61 regqvkgllgdwflcacaalqksrhlldqvppgqpswadqeyrgstcrwqfgrwv 120
QY 121 EVTTDDRLPCLAGLRCFSRCOREDFWLPILLEKVIYAKVHGSYEHHLWAGQVADALVDLTGG 180
Db 121 evttddrlpclagrlcfsrcoredvfwlpillekvyakvhgsyehlhwagqvadalvldltgg 180
QY 181 LAERNLKGAVSGGGQODRPRWEHRTCRQLLHKDOCLISCCVLSPRAGARELGEHAF 240
Db 181 laerwnlkgvagsggqodrpwrhrtcrqlhlkdcclisccvlspragarelgelghaf 240
QY 241 IVSDRLQOAGQCILLRIQNPWGRRCWQWLRGEGGWSQVDAVAASELSQOQEGE 300
Db 241 ivsdrlqoagqcillrlqnpwgrrcwqlwreggwsqvaavaasellsqleqe 300
QY 301 FWEEEFELREFDELTVGYPTVTEAGHLQSLYTERLLCHTRALPGAWKVGOSAGGCRNNSG 360
Db 301 fweeefelrefdelvtgyptvteaghlqsllyterllchtralpawkvkgsgagcrrnns 360
QY 361 FPSNPKFWLRVSEPVSEVYIAVQSRRLHAADWAGRARALVGDSTSWSPASIPGKHQYQAV 420
Db 361 fpsnpkfwlrsvsepvsevyiavlqsrslhaadwagraralvgdstswspasipgkhyqav 420
QY 421 GLHLWKVEKRRVNLPRVLSMPPVAGTACHAYDREHRLCELSPGYIYAVPSTFLKDAPGE 480
Db 421 glhlwkvekrvnlprvlsmpvpvagtachaydrewhrlcelspgyiylavpstflkdapge 480
QY 481 FLFRVSTGRVSLSAIRAVAKNTTPCAALPAGWGTWQWLRGSRVQTAGGSRNFASYPT 540
Db 481 flfrvstgrvslsairavaknttpcaalpawgtwqwlrgsrwvqtaggsrnfasypt 540
QY 541 NPCFPSPVGPGRPCVRITLHQHCRPSDTEFHIFQVPEGGRSODAPPLILOEPL 600
Db 541 npcfpspvgprpcvritlqhcrpsdtefhifqvpeggrrsodappliloep 600
QY 601 LSCVPHRYAQVSRLLCPAGTYKVVPSTYLPDTEGAFVTIATRIDRPSIHSQEMLGOF 660
Db 601 lscvphryaqvsrllcpagtykvvpstylpdtgafvtiatridrpsihsqemlgof 660
QY 661 LQEVSYMAVMK 672
Db 661 lqevsyvmavmk 672

RESULT 4
AAY79575
ID AAY79575 standard; Protein; 666 AA.
AC AAY79575;
XX
XX
DT 15-AUG-2000 (first entry)
DE Mouse calpain 10.
KW Calpain 10; diapaio-1; mouse; diagnosis; therapy; NIDDM1;
KW non-insulin-dependent diabetes mellitus; CAPN10 gene.
XX
OS Mus musculus.
XX
XX WO200023603-A2.
XX
PD 27-APR-2000.
XX
PF 21-OCT-1999; 99WO-US24890.
XX
PR 21-OCT-1998; 98US-0105052.
PR 13-MAY-1999; 99US-0134175.
PA (ARCH-) ARCH DEV CORP.
XX
PI Polonsky KS, Horikawa Y, Oda N, Sreenan S, Zhou Y, Otani K;

PI Hanis CL, Bell GL;
XX
DR WPI: 2000-339702/29.
XX N-PSDB: AA27484, AAY79574, AAY79576.
PT Method for screening for type 2 diabetes mellitus comprises detecting a
PT polymorphism in a calpain encoding nucleic acid segment or a
PT protease-encoding nucleic acid segment -
XX
PS Claim 40; Page 234-236; 257pp; English.
XX
CC The present sequence is that of murine calpain 10 (or diapaio-1), a
CC novel cysteine protease encoded by the major transcript of the
CC CAPN10 gene (see AAA27484). Calpain is ubiquitously expressed in
CC mouse and humans. The invention concerns the identification of
CC genes responsible for type 2 diabetes for use in diagnostic and
CC therapeutic applications. A polymorphism in intron 3 of the human
CC CAPN10 gene (see AAA27475) shows evidence for linkage to type 2
CC diabetes. Claimed methods for screening for a propensity for type
CC 2 diabetes are based on detection of a polymorphism in a calpain
CC encoding nucleic acid. Methods are also claimed for: producing
CC calpain 10 polypeptides using calpain 10-encoding polynucleotides,
CC including the present sequence; identifying modulators of calpain
CC activity using calpain 10 polypeptides and nucleic acids; and using
CC these modulators to treat diabetes, in particular through the
CC regulation of an insulin secretory response or insulin mediated
CC glucose transport.
XX
SQ Sequence 666 AA;

Query Match 81.2%; Score 2963; DB 21; Length 666;
Best Local Similarity 81.1%; Pred. No. 2.le-260;
Matches 544; Conservative 45; Mismatches 76; Indels 6; Gaps 2;

QY 1 MRAGATPAELFRAADSSILFCDLSTPLAQFREDITWRPOETCATPRLPDPDP 60
Db 1 mravtaetparelfrdaafpasdsalfynlstplaqfreditwrpoeicaqpqlfddp 60
QY 61 REGOVKGLLGDWFLCACAALOKSRHLLDVIPPGQPSWADQYRGSTCRWQFGRWV 120
Db 61 regqvkgllgdwflcacaalqksrhlldqvppgqpswadqeyrgstcrwqfghwe 120
QY 121 EVTTDDRLPCLAGLRCFSRCOREDFWLPILLEKVIYAKVHGSYEHHLWAGQVADALVDLTGG 180
Db 121 evttddrlpclagrlcfsrcoredvfwlpillekvyakvhgsyehlhwagqvadalvldtgs 180
QY 181 LAERNLKGAVSGGGQODRPRWEHRTCRQLLHKDOCLISCCVLSPRAGARELGEHAF 240
Db 181 laerwnlkgvagsggqodrpwrhrtcrqlhlkdcclisccvlspragarelgelghaf 240
QY 241 IVSDRLQOAGQCILLRIQNPWGRRCWQWLRGEGGWSQVDAVAASELSQOQEGE 300
Db 241 ivsdrlqoagqcillrlqnpwgrrcwqlwreggwsqvaavaasellsqleqe 300
QY 301 FWEEEFELREFDELTVGYPTVTEAGHLQSLYTERLLCHTRALPGAWKVGOSAGGCRNNSG 360
Db 301 fweeefelrefdelvtgyptvteaghlqsllyterllchtralpawkvkgsgagcrrnsc 360
QY 361 FPSNPKFWLRVSEPVSEVYIAVQSRRLHAADWAGRARALVGDSTSWSPASIPGKHQYQAV 420
Db 361 fpsnpkfwlrsvsepvsevyiavlqsrslhaadwagraralvgdstswspasipgkdyqav 414
QY 421 GLHLWKVEKRRVNLPRVLSMPPVAGTACHAYDREHRLCELSPGYIYAVPSTFLKDAPGE 480
Db 415 glhlwkvekrvnlprvlsmpvpvagtachaydrewhrlcelspgyiylavpstflkdpgq 474
QY 481 FLFRVSTGRVSLSAIRAVAKNTTPCAALPAGWGTWQWLRGSRVQTAGGSRNFASYPT 540
Db 475 flfrvftgklsavrlatkgaspqcalpagewetvqlgdgcrwagtaggsrnfasypt 534
QY 541 NPCFPSPVGPGRPCVRITLHQHCRPSDTEFHIFQVPEGGRSODAPPLILOEPL 600
Db 541 npcfpspvgprpcvritlqhcrpsdtefhifqvpeggrrsodappliloep 600

DB	Accession	Protein Name	Location/Qualifiers
Db	535	ncplrfsvpeagsgpryiriltlqghorltsdqlhnpdgltvfvypddgndacslillqep1 554	
QY	601	LSGVPRHRAQESRLCLLPAGTYKAVPESTYIPDTGAFVTIATRIDPSIHSGEM/GCF 660	
Db	595	lscvphtrvyaqevsrilcllsvngyrlvpecltvlrtegtfvrlatrdqslhsgemlqpl 654	
QY	661	IQEVSVMAYMK 671	
Db	655	IQEVSVMAYMK 665	
RESULT 5			
ID	AAE09756	standard; Protein: 517 AA.	
XX	AAE09756;		
XX	29-NOV-2001	(first entry)	
DE	Human calpain-like protease, 18036.		
KM	Calpain-like protease 18036; spleen disorder; splenomegaly; lung;		
KM	adult respiratory distress syndrome; colon; liver; jaundice; brain;		
KM	idiopathic inflammatory bowel disease; cerebrovascular disease;		
KM	acute meningitis; Alzheimer's disease; T-cell; transplant rejection;		
KM	systemic lupus erythematosus; skin; seborrheic keratosis; leukaemia;		
KM	haematopoietic stem cell; heart; myocardial infarction; atherosclerosis;		
KM	kidney; acute proliferative glomerulonephritis; urolithiasis; apoptosis;		
KM	ischemia; neurodegenerative disease; demyelinating disease; injury; BAE		
KM	LGMD2A muscular dystrophy; proliferative disorder; cancer; gene therapy;		
XX	human; chromosome 2.		
OS	Homo sapiens.		
Key	Location/Qualifiers		
FT	6..11		
FT	/note- "N-myristoylation site"		
FT	7..329		
FT	/note- "Cysteine protease domain"		
FT	13..322		
FT	/label- Calpain_domain		
FT	42..44		
FT	/note- "Protein kinase C phosphorylation site"		
FT	52..54		
FT	/note- "Protein kinase C phosphorylation site"		
FT	67..78		
FT	/note- "Bakaryotic thiol (cysteine) proteases		
FT	99..102		
FT	/note- "Caealn kinase II phosphorylation site"		
FT	107..112		
FT	/note- "N-myristoylation site. The specification		
FT	states that residues GARPAR are present in this site.		
FT	However residues GGFTR are present in the sequence		
FT	given in the specification"		
FT	110..112		
FT	/note- "Protein kinase C phosphorylation site"		
FT	123..126		
FT	/note- "Casein kinase II phosphorylation site"		
FT	189..194		
FT	/note- "N-myristoylation site"		
FT	207..209		
FT	/note- "Protein kinase C phosphorylation site"		
FT	226..228		
FT	/note- "Protein kinase C phosphorylation site"		
FT	265..268		
FT	/note- "amidation site"		
FT	282..285		
FT	/note- "Casein kinase II phosphorylation site"		
FT	332..334		
FT	/note- "Protein kinase C phosphorylation site"		
FT	338..490		
Domain			

FT	Modified-site	/note-"Calpain large subunit domain"
FT	349..354	
FT	/note-"N-myristoylation site"	
FT	357..360	
FT	Modified-site	/note-"N-glycosylation site"
FT	467..469	
FT	Modified-site	/note-"Protein kinase C phosphorylation site"
FT	489..492	
FT	/note-"Casein kinase II phosphorylation site"	
PN	WO200164919-A2.	
PD	07-SEP-2001.	
XX		
PF	26-FEB-2001: 2001WO-US06110.	
XX		
PR	28-FEB-2000: 2000US-0185333.	
XX		
PA	(MILL-) MILLENNIUM PHARM INC.	
XX		
PI	Kapeller-Libermann R;	
XX		
DR	WPI: 2001-565507/63.	
XX	N-PSDB; AAO13883.	
PT		
PT	New calpain-like protease polypeptides and polynucleotides for	
PT	diagnosing, treating seven-transmembrane protein/receptor-related	
PT	disorders and to identify modulators of therapeutic use	
PS	Claim 9; Fig 1; 114pp: English.	
XX		
XX	The invention relates to polypeptide and polynucleotide of calpain-like	
CC	protease, 18036. The polypeptide and polynucleotide of the invention are	
CC	useful for treatment and diagnosis of calpain-like protease-mediated	
CC	disorders. These disorders include disorders of spleen (e.g.	
CC	splenomegaly), lung (e.g. adult respiratory distress syndrome), colon	
CC	(e.g. idiopathic inflammatory bowel disease), liver (e.g. jaundice),	
CC	brain (e.g. traumatic brain injury, cerebral oedema, cerebrovascular	
CC	disease, acute meningitis, Alzheimer's disease), T-cells (e.g. transplant	
CC	rejection, systemic lupus erythematosus), skin (e.g. seborrheic	
CC	keratosis), haematopoietic stem cells (e.g. leukaemias), heart (e.g.	
CC	myocardial infarction, atherosclerosis), kidney (e.g. acute proliferative	
CC	glomerulonephritis, urolithiasis), thymus, breast, testis, uterus,	
CC	epididymis, endometrium, prostate, thyroid, skeletal muscle, pancreas,	
CC	small intestine, disorders related to reduced platelet number, bone,	
CC	ovary, pain and infectious disorders. The molecules of the invention are	
CC	also useful for diagnosing and treating disorders associated with	
CC	perturbed cellular growth and differentiation, exercise-induced injury	
CC	and repair, apoptosis, ischaemia, neurodegenerative diseases,	
CC	demyelinating diseases including experimental allergic encephalomyelitis	
CC	(EAE) and multiple sclerosis, LGMD2A muscular dystrophy, spinal cord	
CC	injury, proliferative and differentiative disorders e.g. cancer and	
CC	renal cell death associated with diverse toxicants. The present sequence	
CC	is human calpain-like protease, 18036. The 18036 gene is located on	
CC	chromosome 2 between D2S140 and D2S2338.	
XX		
Sequence	517 AA:	
50		
Query Match	74.5%: Score 2719.5; DB 22; Length 517;	
Best Local Similarity	76.9%: Pred. No. 2.1e-238;	
Matches	517; Conservative 0; Mismatches 0; Indels 155; Gaps 1	
0Y	1	NRAGSGATPARELFRDAAPPAADSLDFCSLSPPLAOFREDITRRRQETCATRPLPPDP 60
Db	1	mragsaghparelfrdaatpaadsstfcdslspplagfreditwrrpqelcatrplppdp 60
0Y	61	RRGQVKGSLGSCWFLCACAAIQKSNHLDDVYLPQGPSWADQRYGSGFTCRIMGRRV 120
Db	61	regqvkqgllgdcwflcacaalqkshllldvylpqgpswadqeyrgsfctirwfgyrv 120
0Y	121	EYTPDRLRPLCAGRLCSFCOREDEYFWLPLLEVVAKVGVSEPHIMAGQADALVDTGCG 180

Db 121 evttddrlpclagrlcfscrcqredvfwlpillekvaykvhysyehlwagqvadavldltgg 180
Qy 181 LAERNLKGAVAGSGQDRGGRWHEHRTROLLHLKDCCLISCCVLSPRAGARELGEFHAF 240
Db 181 laerwnlkgvagsggqdrpgrwhehrtcrqlhlhkdqcllscvlspragarelgfhaf 240
Qy 241 IVSDRELOQAGOCILLRIQNPWGRWGLWREGGEGWSQVDAVAASELLSQLEGE 300
Db 241 ivsdrlqggagqclllrlqnpwgrwglwreggegswqdaavaseillsqleqe 300
Qy 301 FWEEEFRLREFDELTVGYPTVTEAGHLQSLYTERLLCHTRALPGAWKVGOSAGGCRNNSG 360
Db 301 fweeefrlrefdelvtgypvteaghlqsllyterllchtralpawwkgvagsaggrcnns 360
Qy 361 FPSNPKFWLRVSESEVYIAVLRSLHAADWAGRARALVGDSTWSPASIPGKHQYAV 420
Db 361 fpsnpgfwlrsvsepsvyyiavlqsrllhaadwagraralgdshtswspasipgkhyqav 420
Qy 421 GLHLWKVEKRRVNLPRVLSMPPVAGTACHAYDREVHLRCLESPGYILAVSTFLKADPGE 480
Db 421 ghlwkw ----- 426
Qy 481 FLRLVFTGVRSLSAIRAVAKNTTPGAALPAGEWGTVOLRGSRVWGOTAGSRNFASYPT 540
Db 427 ----- 426
Qy 541 NPCFFSVPEGPGRCVRITLHQHCRPSDTEFHPHIGFHVQVPEGGRSODAPPLLQLEPL 600
Db 427 ----- 426
Qy 601 LSCVPHRYAQVSRVLCCLLPAGTKVVPSTYLPDTEGAFVTIATRIDRPSIHSEMIQGF 660
Db 446 lscvphryaqvserlcllpagtkvvpstypdpdegafvtiatridrpsihsqemlqgf 505
Qy 661 LQEVSMVAVNKT 672
Db 506 lqevsmavnkt 517
RESULT 6
ID AAY79569
XX AAY79569 standard; Protein; 517 AA.
AC AAY79569;
XX 15-AUG-2000 (first entry)
XX Human calpain 10c.
XX /NIDM1; non-insulin-dependent diabetes mellitus; CAPN10 gene;
KW calpain 10; calpain 10c; diapain-1; cysteine protease; diagnosis;
KW therapy.
XX Homo sapiens.
XX WO200023603-A2.
XX 27-APR-2000.
XX 21-OCT-1999; 99WO-US24890.
XX 21-OCT-1998; 98US-0105052.
XX 13-MAY-1999; 99US-0134175.
XX (ARCH-) ARCH DEV CORP.
XX Polonsky KS, Horikawa Y, Oda N, Sreenan S, Zhou Y, Otani K;
PI Hanis CL, Bell GI;
XX WPI: 2000-339702/29.
XX N-PSDB; AAA27475, AAA27578.
XX Method for screening for type 2 diabetes mellitus comprises detecting a

PT polymorphism in a calpain encoding nucleic acid segment or a
PT protease-encoding nucleic acid segment -
XX Claim 40; Page 223-224; 257pp; English.
XX
CC The present sequence is that of a novel human cysteine protease,
CC designated calpain 10c. The protein is encoded by a transcript
CC (see AAA27478) of the human calpain 10 gene, CAPN10 (see AAA27475),
CC which has been identified in the NIDDM1 region of chromosome 2.
CC The CAPN10 gene consists of 15 exons spanning 32 kb. A complex
CC pattern of alternative splicing generates proteins of 672, 544,
CC 517, 513, 444, 274, 139 and 138 amino acids (see AAY79567-74),
CC designated calpain 10a to 10h, respectively. Calpain 10c is
CC readily detectable in many tissues, including skeletal muscle and
CC islets. The invention concerns the identification of genes
CC responsible for type 2 diabetes for use in diagnostic and
CC therapeutic applications. A G-to-A polymorphism in intron 3 of the
CC CAPN10 gene (UCSNP-43) shows evidence for linkage to type 2
CC diabetes. Claimed methods for screening for a propensity for type
CC 2 diabetes mellitus are based on detection of a polymorphism in a
CC calpain encoding nucleic acid, especially UCSNP-10 of the CAPN10
CC gene. Methods are also claimed for: producing calpain 10
CC polypeptides using calpain 10-encoding polynucleotides; regulating
CC diabetes by providing a calpain polypeptide such as calpain 10c;
CC identifying modulators of calpain activity using calpain 10
CC polypeptides; and using these modulators to treat diabetes, in
CC particular through the regulation of an insulin secretory response
CC or insulin mediated glucose transport.
XX
SQ Sequence 517 AA;
Query Match 74.4%; Score 2715.5; DB 21; Length 517;
Best Local Similarity 76.8%; Pred. No. 4.8e-238;
Matches 516; Conservative 1; Mismatches 0; Indels 155; Gaps 1;
QY 1 MRAGRGATPARELFRDAAPFAADSSLFCDLSTPLAQFREDITWRROELCATPLRPDDP 60
Db 1 mragrgatparelfrdaafpaadsalfcdlstplaqfreditwrpqelcatplrfddp 60
QY 61 REGQVKGILLGDCWFLCACAALQKSRHLDDQVIPPQPSWADQYRGSTCRWQGRWV 120
Db 61 regqvkgllgdcwflcacaalqksrhlldqvippqpswadqyrgstcrwqgrwv 120
QY 121 EVTTDDRLPCLAGRLCFSCRCQREDVFWLPILLEKVYKAVHYSYEHHLWAGVADALVLTGG 180
Db 121 evttddrlpclagrlcfscrcqredvfwlpillekvaykvhysyehlwagvadalvltgg 180
QY 181 LAERNLKGAVAGSGQDRGGRWHEHRTCRQLHLKDCCLISCCVLSPRAGARELGEFHAF 240
Db 181 laerwnlkgvagsggqdrpgrwhehrtcrqlhlhkdqcllscvlspragarelgfhaf 240
QY 241 IVSDRELOQAGOCILLRIQNPWGRWGLWREGGEGWSQVDAVAASELLSQLEGE 300
Db 241 ivsdrlqggagqclllrlqnpwgrwglwreggegswqdaavaseillsqleqe 300
QY 301 FWEEEFRLREFDELTVGYPTVTEAGHLQSLYTERLLCHTRALPGAWKVGOSAGGCRNNSG 360
Db 301 fweeefrlrefdelvtgypvteaghlqsllyterllchtralpawwkgvagsaggrcnns 360
QY 361 FPSNPKFWLRVSESEVYIAVLRSLHAADWAGRARALVGDSTWSPASIPGKHQYAV 420
Db 361 fpsnpgfwlrsvsepsvyyiavlqsrllhaadwagraralgdshtswspasipgkhyqav 420
QY 421 GLHLWKVEKRRVNLPRVLSMPPVAGTACHAYDREVHLRCLESPGYILAVSTFLKADPGE 480
Db 421 ghlwkw ----- 426
QY 481 FLRLVFTGVRSLSAIRAVAKNTTPGAALPAGEWGTVOLRGSRVWGOTAGSRNFASYPT 540
Db 427 ----- 426
QY 541 NPCFFSVPEGPGRCVRITLHQHCRPSDTEFHPHIGFHVQVPEGGRSODAPPLLQLEPL 600


```

Db 427 -----|||
601 LSCVPRHVAQEVSRLLCPAGTKYVPSYTLPTGCAFTVTATRTDRPSHSOENLGP 660
446 LSCVPHRYQEVSRLLCPAGTKYVPSYTLPTGCAFTVTATRTDRPSHSOENLGP 505
661 LQEVSYMAVMKKT 672
506 LQEVSYMAVMKKT 517

RESULT 7
AA79568
ID AA79568 standard; Protein: 544 AA.
AC AA79568;
DE 15-AUG-2000 (first entry)
DE Human calpain 10b.
KW NIDDM1; non-insulin-dependent diabetes mellitus; CAPN10 gene;
KW calpain 10; calpain 10b; diapain-1; cysteine protease; diagnosis;
KW therapy.
OS Homo sapiens.
PN M0200023603-A2.
DE 27-APR-2000.
DE 21-OCT-1999; 99WO-US24890.
DE 21-OCT-1998; 98US-0105052.
DE 13-MAY-1999; 99US-0134175.
PA (ARCH-) ARCH DEV CORP.
PI Polonsky KS, Horikawa Y, Oda N, Sreenan S, Zhou Y, Otsu K;
PI Hanis CL, Bell GI;
DR WPI: 2000-339702/29.
DR N-PSDB: AAA27475, AAA27477.
PT Method for screening for type 2 diabetes mellitus comprises detecting a
PT polymorphism in a calpain encoding nucleic acid segment or a
PT protease-encoding nucleic acid segment -
PS Claim 40; Page 220-222; 257pp; English.
XX The present sequence is that of a novel human cysteine protease,
XX designated calpain 10b. The protein is encoded by a transcript
XX (see AA27477) of the human calpain 10 gene, CAPN10 (see AA27475),
XX which has been identified in the NIDDM1 region of chromosome 2.
XX The CAPN10 gene consists of 15 exons spanning 32 kb. A complex
XX pattern of alternative splicing generates proteins of 672, 544,
XX 517, 513, 444, 274, 139 and 138 amino acids (see AA79567-74),
XX designated calpain 10a to 10h, respectively. Calpain 10b is
XX readily detectable in many tissues, including skeletal muscle and
XX islets. The invention concerns the identification of genes
XX responsible for type 2 diabetes for use in diagnostic and
XX therapeutic applications. A G-to-A polymorphism in Intron 3 of the
XX CAPN10 gene (UCSNP-43) shows evidence for linkage to type 2
XX diabetes. Claimed methods for screening for a propensity for type
XX 2 diabetes mellitus are based on detection of a polymorphism in a
XX calpain encoding nucleic acid, especially UCSNP-10 of the CAPN10
XX gene. Methods are also claimed for: producing calpain 10
XX polypeptides using calpain 10-encoding polynucleotides; regulating
XX diabetes by providing a calpain polypeptide such as calpain 10b;
XX identifying modulators of calpain activity using calpain 10
XX polypeptides; and using these modulators to treat diabetes, in
XX particular through the regulation of an insulin secretory response

```

```

CC or Insulin mediated glucose transport.
XX
SQ Sequence 544 AA:
Query Match 74.0%; Score 2701.5; DB 21; Length 544;
Best Local Similarity 95.4%; Pred. No. 9,6e-237;
Matches 498; Conservative 5; Mismatches 6; Indels 13; Gaps 1;

QY 1 MRAGCATPARELFRDAAPADSSILFCDSLSTPLAOFREDITWRAPQECATPRLPPDP 60
DB 1 MRAGCATPARELFRDAAPADSSILFCDSLSTPLAOFREDITWRAPQECATPRLPPDP 60
QY 61 REGOVKQGLGDCWFLCACALQKSRHLLDOYIPEGQPSWADQEXRGSTCRITWQGRNV 120
DB 61 REGOVKQGLGDCWFLCACALQKSRHLLDOYIPEGQPSWADQEXRGSTCRITWQGRNV 120
QY 121 EYTTDRRLPCLAGRLCFSRGCRDVEWFLPLEKVKYAKVGSYEHIMAGOVADALVDLNGG 180
DB 121 EYTTDRRLPCLAGRLCFSRGCRDVEWFLPLEKVKYAKVGSYEHIMAGOVADALVDLNGG 180
QY 181 LAERNMLKGVAGSGGQODRPGRMENRTCRQLHLKDOCLISCVLSPRAGARELSEFFNAF 240
DB 181 LAERNMLKGVAGSGGQODRPGRMENRTCRQLHLKDOCLISCVLSPRAGARELSEFFNAF 240
QY 241 IYSDRLRLGQAGCCILLRIQNPWGRGRCWGLMBEGGSGQVDANAVASSELLSQLQEGE 300
DB 241 IYSDRLRLGQAGCCILLRIQNPWGRGRCWGLMBEGGSGQVDANAVASSELLSQLQEGE 300
QY 301 FVVEEEFLREDELTVGVPTVEAGHLQSLYERLLCHTRALPGAMVQSGSGGRNNG 360
DB 301 FVVEEEFLREDELTVGVPTVEAGHLQSLYERLLCHTRALPGAMVQSGSGGRNNG 360
QY 361 FPSNPKFWLRSEPESEVYIAVLQSRNLHADAAGARALVGSHTSWSPASIPGRHYQAV 420
DB 361 FPSNPKFWLRSEPESEVYIAVLQSRNLHADAAGARALVGSHTSWSPASIPGRHYQAV 420
QY 421 GLHMKVEKRRNLPRLVLSMPVYAGTACHAYREYHLKCELSPGYYLAVSTFLDAPGE 480
DB 421 GLHMKVEKRRNLPRLVLSMPVYAGTACHAYREYHLKCELSPGYYLAVSTFLDAPGE 480
QY 481 FLRVSTGRVLSAIVAAK-----NTTPGAAL 509
DB 481 FLRVSTGRVLSAIVAAK-----NTTPGAAL 509

RESULT 8
AA79570
ID AA79570 standard; Protein: 513 AA.
AC AA79570;
DE 15-AUG-2000 (first entry)
DE Human calpain 10d.
KW NIDDM1; non-insulin-dependent diabetes mellitus; CAPN10 gene;
KW calpain 10; calpain 10d; diapain-1; cysteine protease; diagnosis;
KW therapy.
OS Homo sapiens.
PN M0200023603-A2.
DE 27-APR-2000.
DE 21-OCT-1999; 99WO-US24890.
DE 21-OCT-1998; 98US-0105052.
DE 13-MAY-1999; 99US-0134175.
PA (ARCH-) ARCH DEV CORP.
XX

```

PI Polonsky KS, Horikawa Y, Oda N, Sreenan S, Zhou Y, Otani K;
PI Hanis CL, Bell GI;
XX WPI; 2000-339702/29.
DR N-PSDB; AAA27475, AAA27479.
DR
XX
PT Method for screening for type 2 diabetes mellitus comprises detecting a
PT polymorphism in a calpain encoding nucleic acid segment or a
PT protease-encoding nucleic acid segment -
XX
PS
PS
XX
XX
CC The present sequence is that of a novel human cysteine protease,
CC designated calpain 10d. The protein is encoded by a transcript
CC (see AAA27479) of the human calpain 10 gene, CAPN10 (see AAZ27475),
CC which has been identified in the NIDDM1 region of chromosome 2.
CC The CAPN10 gene consists of 15 exons spanning 32 kb. A complex
CC pattern of alternative splicing generates proteins of 672, 544,
CC 517, 513, 444, 274, 139 and 138 amino acids (see AAY79567-74),
CC designated calpain 10a to 10h, respectively. Calpain 10d is
CC expressed only at low levels in the tissues examined. The
CC invention concerns the identification of genes responsible for type
CC 2 diabetes for use in diagnostic and therapeutic applications. A
CC G-to-A polymorphism in intron 3 of the CAPN10 gene (UCSNP-43) shows
CC evidence for linkage to type 2 diabetes. Claimed methods for
CC screening for a propensity for type 2 diabetes mellitus are based
CC on detection of a polymorphism in a calpain encoding nucleic acid,
CC especially UCSNP-10 of the CAPN10 gene. Methods are also claimed
CC for: producing calpain 10 polypeptides using calpain 10-encoding
CC polynucleotides; regulating diabetes by providing a calpain
CC polypeptide such as calpain 10d; identifying modulators of calpain
CC activity using calpain 10 polypeptides; and using these modulators
CC to treat diabetes, in particular through the regulation of an
CC insulin secretory response or insulin mediated glucose transport.
XX
SQ Sequence 513 AA;

Query Match 74.0%; Score 2699; DB 21; Length 513;
Best Local Similarity 98.2%; Pred. No. 1.5e-236;
Matches 496; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 1 MRAGRGATPARELFDAAPFAADSSFLCDLSTPLAQREDITWRPOEICATRLFPDDP 60
Db 1 mragrgatparelfdaapfaadssflcdlstplagfreditwrrpoeicatrlfpddp 60

QY 61 REGQVQGLIGDCWFLCACAAKSRHLLDQVIPPQGPSWADQYRGSTCRWQGRWV 120
Db 61 regqvqgligdcwflcacaaakshrlldqvipppqgpswadqeyrgstcrwqgrwv 120

QY 121 EVTTDRPLCLAGRLCFSCQREDVFWLPLEKRYAKVHGSYEHLMWAGQVADALVLTGG 180
Db 121 evttddrlpclagrlcfscrqredvfwlplekryakvghsyehlmwagqvadalvltgg 180

QY 181 LAERNLKGVAGSGGODRPGREHETCRQLLHLKDKQCLISCCVLSFRAGRELGEFAP 240
Db 181 laernlkgvagsggodrpgrwehrtcrqlhlhkdqclisccvlsfragarelgefap 240

QY 241 IVSDLRELQAGQACILLRIQNPWGRRCWGLWRGEGGWSOVDAVASSELLSQLEGE 300
Db 241 ivsdhrelqagqacillriqnpwgrrcwglwrgeggsowdaavasellsqleqe 300

QY 301 FWVEEFDELJVGVPVTEAGHLQSLYTERLCHTRALPGAWKVGOSAGGCRNNSG 360
Db 301 fwveeefdeljvgvpvteaghlqsllyterlchtralgawvkvgosaggcrnnsg 360

QY 361 FPNRPFWLRVSPSEVYTAVALORSRLHAADWAGRARALVGDSTSWSPASIPGKHQAV 420
Db 361 fspnrfwlrvspsevytavalorsrlhaadwagraralvgdstswspasipgkhqav 420

QY 421 GLHLKVKERRRNLPRVLSMPVAGTACHAYDREHLRCELSPGYLYAVPSTFLKDAPGE 480
Db 421 glhlkwkerrrnlprvlsmppvagtachaydrehlrcelsgpylyavpstflkdapge 480

QY 481 FLLRVFSTGRVSLSAIRAVAKNTTP 505
Db 481 flrvfstgrvslsairavaknttp 505

RESULT 9

AAV79571
ID AAY79571 standard; Protein: 444 AA.

XX
AC AAY79571;

XX
DT 15-AUG-2000 (first entry)

XX
DE Human calpain 10e.

XX
KW NIDDM1; non-insulin-dependent diabetes mellitus; CAPN10 gene;
KW calpain 10; calpain 10e; diapain-1; cysteine protease; diagnosis;
KW therapy.

XX
OS Homo sapiens.

XX
PN WO200023603-A2.

XX
PD 27-APR-2000.

XX
PF 21-OCT-1999; 99WO-US24890.

XX
PR 21-OCT-1998; 98US-0105052.

XX
PR 13-MAY-1999; 99US-0134175.

XX
PA (ARCH-) ARCH DEV CORP.

XX
PI Polonsky KS, Horikawa Y, Oda N, Sreenan S, Zhou Y, Otani K;

PI Hanis CL, Bell GI;

XX
WPI; 2000-339702/29.

DR N-PSDB; AAA27475, AAA27480.

DR Method for screening for type 2 diabetes mellitus comprises detecting a
DR polymorphism in a calpain encoding nucleic acid segment or a
DR protease-encoding nucleic acid segment -

XX
Claim 40; Page 228-229; 257pp; English.

XX The present sequence is that of a novel human cysteine protease,
XX designated calpain 10e. The protein is encoded by a transcript
XX (see AAA27480) of the human calpain 10 gene, CAPN10 (see AAZ27475),
XX which has been identified in the NIDDM1 region of chromosome 2.
XX The CAPN10 gene consists of 15 exons spanning 32 kb. A complex
XX pattern of alternative splicing generates proteins of 672, 544,
XX 517, 513, 444, 274, 139 and 138 amino acids (see AAY79567-74),
XX designated calpain 10a to 10h, respectively. Calpain 10e is
XX expressed only at low levels in the tissues examined. The
XX invention concerns the identification of genes responsible for type
XX 2 diabetes for use in diagnostic and therapeutic applications. A
XX G-to-A polymorphism in intron 3 of the CAPN10 gene (UCSNP-43) shows
XX evidence for linkage to type 2 diabetes. Claimed methods for
XX screening for a propensity for type 2 diabetes mellitus are based
XX on detection of a polymorphism in a calpain encoding nucleic acid,
XX especially UCSNP-10 of the CAPN10 gene. Methods are also claimed
XX for: producing calpain 10 polypeptides using calpain 10-encoding
XX polynucleotides; regulating diabetes by providing a calpain
XX polypeptide such as calpain 10e; identifying modulators of calpain
XX activity using calpain 10 polypeptides; and using these modulators
XX to treat diabetes, in particular through the regulation of an
XX insulin secretory response or insulin mediated glucose transport.

XX
SQ Sequence 444 AA;

Query Match 64.0%; Score 2336; DB 21; Length 444;
Best Local Similarity 99.8%; Pred. No. 1.2e-203;

Matches 425; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MRAGCATPARELFPAAPADSSLFCDLSPFAOFREDIMRRROELCAPRLEPPDP 60
 Db 1 mragcatparelfrdaaipaasllfcdlspfaofredlfrwrrqelcaprlfppdp 60
 OY 61 REGVYKQGLLGDQWFLCACAALQKSRHLDOVIPPQPSMADQERYGFTCRIMQGRWV 120
 Db 61 regvykqglldgcwflcacaalqkshlldqvippqpswadqerygftcrimqgrwv 120
 OY 121 EYTTDDRLPCLAGRLCFSCRCOREVFWPLLEKYAVAKVGSYEHLMAGVADALVDLTG 180
 Db 121 eyttddrlpclagrlcfscrcorevfwplllekyavakvgsyehlmagvadaldvltg 180
 OY 181 LAERNMLKCVAGSGGQODRGRMEHRTCRQLHLKDOCLISCCVLSFRAGARELSEFNAF 240
 Db 181 laernmlkcvagsggqodrpgrmehrtcrqlhlkdoclisccvlsfragarelsefna 240
 OY 241 IVSDIRELQAGQOCITLLRIQNPWGRRCMOGLMREGSGMSQVDAVAASELSQLEGE 300
 Db 241 ivsdirelqagqocitllriqnpwgrrcmoglmregsgmsqvdaavaaselsglege 300
 OY 301 FVWESEERFREDLTGVPVTEAGHLSLYTERLLCHTRALPGAWVKQSGAGGRNNSG 360
 Db 301 fvweeseerfiredltvgvpteaghlsglyterllchtralpawvkgsgaggrnns 360
 OY 361 FFSNRFMLRVSESEVYAVQBSRIHAADAGARALVGSYSHSWSASIPGKHQYAV 420
 Db 361 ffsnrfmlrvsesevayavqbsrihaadagarralvgshshtswspasipghyqav 420
 OY 421 GHLHMK 426
 Db 421 ghhlvk 426

RESULT 10

AAM25421
 ID AAM25421 standard; Protein: 245 AA.

AC AAM25421;

DT 16-OCT-2001 (first entry)

DE Human protein sequence SEQ ID NO:936.

XX Human: cancer; ulcer; HIV infection; human immunodeficiency virus;
 KW antiinflammatory; antirheumatic; antiallergic; immunosuppressive;
 KW antibacterial; endocrine; cardiac; central nervous system; vitruce;
 KW anti-HIV; fungicide; antitumour; cardiovascular; antianemic; anemia;
 KW antiagregant; haemostatic; vulnary; antileuk; osteopathic; eczema;
 KW dermatological; antiallergic; antidiabetic; antidiabetic; cytostatic;
 KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
 KW immunostimulant; gene therapy; antitense therapy; vaccine; inflammation;
 KW antiinflammatory; rheumatoid arthritis; septic shock; pancreatitis;
 KW cardiac dysfunction; neuropathic disorder; cardiac anaphylaxis; autoimmunity;
 KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
 KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
 KW allergic rhinitis; diabetes; multiple sclerosis; depression;
 KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
 KW neurological disorder.

XX Homo sapiens.

XX W0200153455-A2.

XX 26-JUL-2001.

XX 22-DEC-2000; 2000WO-US35017.

XX 23-DEC-1999; 99US-0471275.

XX 21-JAN-2000; 2000US-0488725.
 XX 25-APR-2000; 2000US-0552317.

PA (HYSE-) HYSEQ INC.

XX Tang YF, Liu C, Dimauc RT;

XX WPI, 2001-457603/49.

DR N-PSDB: AAH99362.

PT Isolated human polynucleotides encoding polypeptides, useful for the treatment and diagnosis of e.g. cancer, ulcers and HIV infection -

PS Claim 20; Page 201; 1217p; English.

CC AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
 CC AAM25963. The proteins can have activities based on the tissues and
 CC cells they are expressed in, such as: antiinflammatory; antirheumatic;
 CC antiallergic; immunosuppressive; antibacterial; endocrine; cardiac;
 CC central nervous system; vitruce; anti-HIV; fungicide; antitumour;
 CC cardiovascular; antianemic; antiagregant; haemostatic; vulnary;
 CC antileuk; osteopathic; dermatological; antiallergic; antidiabetic;
 CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
 CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
 CC encoding them can be used in gene therapy, antisense therapy and vaccine
 CC production. The proteins and polynucleotides are useful for screening for
 CC agonists or antagonists of a protein and for the treatment and diagnosis
 CC of disorders associated with the activity of a protein e.g. inflammation,
 CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
 CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
 CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
 CC anemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
 CC osteoporosis, severe combined immunodeficiency, eczema, allergic
 CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
 CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
 CC neurological disorders.

CC Sequence 245 AA;

Query Match 26.4%; Score 964; DB 22; Length 245;
 Best Local Similarity 89.7%; Pred. No. 3.2e-79;
 Matches 175; Conservative 2; Mismatches 2; Indels 16; Gaps 1;

OY 60 PREGVYKQGLLGDQWFLCACAALQKSRHLDOVIPPQPSMADQERYGFTCRIMQGRW 119
 Db 1 pregvykqglldgcwflcacaalqkshlldqvippqpswadqerygftcrimqgrw 60

OY 120 VEYTTDDRLPCLAGRLCFSCRCOREVFWPLLEKYAVAKVGSYEHLMAGVADALVDLTG 179
 Db 61 veyttddrlpclagrlcfscrcorevfwplllekyavakvgsyehlmagvadaldvltg 120

OY 180 GLAERNMLKCVAGSGGQODRGRMEHRTCRQLHLKDOCLISCCVLSFRAGARELSEFNA 239
 Db 121 glaernmlkcvagsggqodrpgrmehrtcrqlhlkdoclisccvlsfragarelsefna 171

OY 240 FIVSDIRELQAGQOCITLLRIQNPWGRRCMOGLMREGSGMSQVDAVAASELSQLEGE 254
 Db 172 fivsdirelqagqocitllriqnpwgrrcmoglmregsgmsqvdaavaaselsglege 179

RESULT 11

AA79572
 ID AA79572 standard; Protein: 274 AA.

AC AA79572;

DT 15-AUG-2000 (first entry)

DE Human calpain 10f.

XX NIDDM1; non-insulin-dependent diabetes mellitus; CAPN10 gene;
 KW calpain 10; calpain 10f; diapaen-1; cysteine protease; diagnosis;
 KW therapy.

OS Homo sapiens.

XX PN WO200023603-A2.
XX PD 27-APR-2000.
XX PF 21-OCT-1999; 99WO-US24890.
XX PR 21-OCT-1998; 98US-0105052.
XX PR 13-MAY-1999; 99US-0134175.
XX PA (ARCH-) ARCH DEV CORP.
XX PI Polonsky KS, Horikawa Y, Oda N, Sreenan S, Zhou Y, Otani K;
XX PI Hanis CL, Bell GI;
XX DR WPI; 2000-339702/29.
XX DR N-PSDB; AAA27475, AAA27481.
XX PT Method for screening for type 2 diabetes mellitus comprises detecting a
XX PT polymorphism in a calpain encoding nucleic acid segment or a
XX PT protease-encoding nucleic acid segment -
XX PS Claim 40; Page 230-231; 257pp; English.
XX CC The present sequence is that of a novel human cysteine protease,
XX CC designated calpain 10f. The protein is encoded by a transcript
XX CC (see AAA27481) of the human calpain 10 gene, CAPN10 (see AA227475),
XX CC which has been identified in the NIDDM1 region of chromosome 2.
XX CC The CAPN10 gene consists of 15 exons spanning 32 kb. A complex
XX CC pattern of alternative splicing generates proteins of 672, 544,
XX CC 517, 513, 444, 274, 139 and 138 amino acids (see AA279567-74),
XX CC designated calpain 10a to 10h, respectively. Calpain 10f is
XX CC expressed only at low levels in the tissues examined. The
XX CC invention concerns the identification of genes responsible for type
XX CC 2 diabetes for use in diagnostic and therapeutic applications. A
XX CC G-to-A polymorphism in intron 3 of the CAPN10 gene (UCSNP-43) shows
XX CC evidence for linkage to type 2 diabetes. Claimed methods for
XX CC screening for a propensity for type 2 diabetes mellitus are based
XX CC on detection of a polymorphism in a calpain encoding nucleic acid,
XX CC especially UCSNP-10 of the CAPN10 gene. Methods are also claimed
XX CC for: producing calpain 10 polypeptides using calpain 10-encoding
XX CC polynucleotides; regulating diabetes by providing a calpain
XX CC polypeptide such as calpain 10f; identifying modulators of calpain
XX CC activity using calpain 10 polypeptides; and using these modulators
XX CC to treat diabetes, in particular through the regulation of an
XX CC insulin secretory response or insulin mediated glucose transport.
XX SQ Sequence 274 AA;

Query Match 23.5%; Score 858.5; DB 21; Length 274;
Best Local Similarity 81.7%; Pred. No. 1.5e-69;
Matches 165; Conservative 1; Mismatches 13; Indels 23; Gaps 3;
QY 1 MRAGRGATPARELFDAAPPAADSSLCFLSTPLAQFREDITWRPQECATPRLFPDDP 60
Db 1 mrgrgatparelfdaapdaadsslfcdlstplagfreditwrrpqecatprlfpddp 60
QY 61 REGGVKGLGDCWFLCACAALQKSRHLDPVPPGQPSWADQYRGSTFCRTWQFGRW 120
Db 61 reggvkgllgdcwflcacaalksrhldqvppgqpswadqeyrgstfcrtwqfgrw 120
QY 121 EVTTDRPLCLAGRLCFSCQREDVFWLPLEKVAHVGSYHLWAGOVADALVLTGG 180
Db 121 evttddrlpclagrlcfscqredvfwlplekvaahvgsyhlwagovadalvltgg 180
QY 181 LAERWNLKGVAGSGGQDDPRGR 202
Db 168 -----gcpggdpdrpgr 179

RESULT 12
AAB62155

ID XX AAB62155 standard; Protein; 686 AA.
XX AC AAB62155;
XX DT 29-MAY-2001 (first entry)
XX DE Novel human protein (NHP) #4.
XX KW Novel human protein; NHP; calcium; protease; gene therapy; screening.
XX OS Homo sapiens.
XX PN WO2000116336-A1.
XX PD 08-MAR-2001.
XX PF 01-SEP-2000; 2000WO-US24062.
XX PR 02-SEP-1999; 99US-0152057.
XX PA (LEXI-) LEXICON GENETICS INC.
XX PI Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
XX DR WPI; 2001-235112/24.
XX DR N-PSDB; AAF57305.
XX PT Novel human polynucleotides isolated from human testis cDNA library,
XX PT encodes novel human proteins, useful as reagents in screening for
XX PT compounds used for treating mental and biological disorders -
XX PS Claim 8; Page 33-35; 39pp; English.
XX CC The invention relates to novel human polynucleotides encoding protein
XX CC (NHP) that share sequence similarity with human calcium dependant
XX CC proteases. Nucleotide constructs encoding functional NHPs are used in
XX CC gene therapy for the modulation of NHP expression. NHP oligonucleotides
XX CC can be used as hybridization probes for screening libraries and assessing
XX CC NHP gene expression patterns. The sequences may also be used as part of
XX CC ribozyme and/or triple helix sequences that are useful for NHP gene
XX CC regulation. Labeled NHP nucleotide probes can be used to screen a human
XX CC genomic library. The NHP nucleotide sequences are also useful in drug
XX CC screening techniques. Nucleotide constructs encoding NHP products can be
XX CC used to genetically engineer host cells to express NHP products in vivo,
XX CC these genetically engineered cells function as bioreactors in the body,
XX CC delivering a continuous supply of a NHP, a NHP peptide, or a NHP fusion
XX CC protein to the body. The present sequence represents a NHP.
XX SQ Sequence 686 AA;

Query Match 17.7%; Score 644.5; DB 22; Length 686;
Best Local Similarity 33.7%; Pred. No. 1.7e-49;
Matches 168; Conservative 71; Mismatches 201; Indels 59; Gaps 15;
QY 12 ELFRDAAPPAADSSLCFLSTPLAQFREDITWRPQECATPRLFPDDPREGVQKGLL 70
Db 41 elfrdaapppaadsllfcldstplaqfreditwrrpqecatprlfpddprrgqvkgll 70
QY 71 GDCWFLCACAALQKSRHLDPVPPGQPSWADQYRGSTFCRTWQFGRWVETDRLPC 130
Db 100 gdcwllaaaglsittcpklllyrvvrrgqs--fknyagifhfgwqfgrwvrvvddript 157
QY 131 LAGRLCFSCQREDVFWLPLEKVAHVGSYHLWAGOVADALVLTGGLAERWNLKGV 190
Db 158 kndklvfvfstersefwlsallekayaklsgsyalsggstmgledftggvagsfqlq-- 215
QY 191 AGSGGQDDPRGRWEHRTCPQLLEKDCQLISCCVLSPRAGRE-----LGEFHATVSD 244
Db 216 -----rppqnlrlirka--verssmgcsievtssdslesmtdkmlvgrghaysvtg 265
QY 245 LRELOQAGQCILLRLRIONPWGRRRCWQGLWREGGEGCSOVDAVAASELLSOLQEGEFWE 304
XX SQ

Db 266 lqdvh-yrgkmetllirvnpwgrlewgawdsarewevasdigmqlhktedgefms 324
 QY 305 EEEFLREPELTVG--YPTVAGHLQSLYTERLLCHTRALPGAWVGOSAGCCRRNSG-F 361
 Db 325 yqdlfnftlleicnlptdltsgdyksyw-----htffegswrrsgsgcgrnpgtlf 378
 QY 362 PSNRFKWLRYSE---PSEVYIAVLQSRRLHADWAGR---ARALVGDSTHSMPASIPGK 415
 Db 379 wtnpqfkfklspgddpded-----daegnvvctclvalmqkmwrhargga 424
 QY 416 HYOAVGLHLMKVKKRRVNLPRV-LSMPPVACTACHAY-----DREYHLRCELSPGYTLA 468
 Db 425 qlqtlgfvlyavpketfqnldvnlkkekftkyqdhgfselftnsrevssqrlrppgey11 484
 QY 469 VPSTFLKDAPEGELLRVFS 487
 Db 485 lpsltephrdaadflirvft 503

RESULT 13

AAB40701

ID AAB40701 standard; Protein; 702 AA.

AC AAB40701;

DT 08-FEB-2001 (first entry)

XX Human ORFX ORF465 polypeptide sequence SEQ ID NO:930.

XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 XX vulnary; antiparasitic; antiparkinsonian; noctropic; neuroprotective;
 XX anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 XX immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 XX hypotensive; dermatological; immunosuppressive; antiinflammatory;
 XX antiviral; antibacterial; antifungal; antihemmatic; antihydroid;
 XX antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 XX neurodegenerative disorder; osteoarthritis; graft vs host disease;
 XX cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 XX cholesterol ester storage; systemic lupus erythematosus; infection;
 XX severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 XX allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 XX bone damage; cartilage damage; antiinflammatory disease; coagulation;
 XX thrombosis; contraceptive.

OS Homo sapiens.

PN WO200058473-A2.

XX 05-OCT-2000.

PF 31-MAR-2000; 2000MO-US08621.

XX 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.

PR 05-APR-1999; 99US-0127728.

PR 30-MAR-2000; 2000US-0540763.

XX (CURA-) CURAGER CORP.

XX Shinketsu RA, Leach M;

XX WPI; 2000-602362/57.

DR N-PSDB; AAC74910.

XX Novel nucleic acids and peptides derived from open reading frame X,

XX useful for treating e.g. cancers, proliferative disorders, asthima,

XX neurodegenerative disorders and cardiovascular disease -

CC antipsoriatic; antiparkinsonian; noctropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antineumatic;
 CC antihydroid; and antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

SQ Sequence 702 AA;

Query Match

Best Local Similarity 33.7%; Score 644.5; DB 21; Length 702;

Matches 168; Conservative 71; Mismatches 201; Indels 59; Gaps 15;

QY 12 ELFRDAFPADSSL-FCDISTFLAOPREDITRRRPOEICATPRLFPDDREGOVKGL 70
 Db 41 elfedrlfpepslsjgkldy-pnshvngpntsqwtkdlnlplfmdgispdldcg11 99
 QY 71 GDCWFCAALQKSRHLDDQVTPPGPSWADYRGSFCTCRIMQGRWYVTTDRIPC 130
 Db 100 gdcwllaalsltcpllllyrvprgs--fknyagi fhnfglwqfgrvnnvddrlpt 157
 QY 131 LAGRLFCRQRQRDVFWLPLEVYAKVNGSYELHAGVADLVLTGSLAERNLKG 190
 Db 158 knkdlrvnstersewtsallekayaklsysyalsgs tmeqledftgvgasfqlg-- 215
 QY 191 AGSGGODRPRGWEHNTCRQLHLKQDCVLSPPRAGARE-----LGEFHAFTVSD 244
 Db 216 -----ppqqlhlllrltra--versslngcstevsdelscmtdkmlvrgnhsyvg 265
 QY 245 LRELQAGOCITLLRIONPWRRCWQGLMRGEGGNSOVDAVASELLSQLQEGEFWE 304
 Db 266 lqdvh-yrgkmetllirvnpwgrlewgawdsarewevasdigmqlhktedgefms 324
 QY 305 EEEFLREPELTVG--YPTVAGHLQSLYTERLLCHTRALPGAWVGOSAGCCRRNSG-F 361
 Db 325 yqdlfnftlleicnlptdltsgdyksyw-----htffegswrrsgsgcgrnpgtlf 378
 QY 362 PSNRFKWLRYSE---PSEVYIAVLQSRRLHADWAGR---ARALVGDSTHSMPASIPGK 415
 Db 379 wtnpqfkfklspgddpded-----daegnvvctclvalmqkmwrhargga 424
 QY 416 HYOAVGLHLMKVKKRRVNLPRV-LSMPPVACTACHAY-----DREYHLRCELSPGYTLA 468
 Db 425 qlqtlgfvlyavpketfqnldvnlkkekftkyqdhgfselftnsrevssqrlrppgey11 484
 QY 469 VPSTFLKDAPEGELLRVFS 487
 Db 485 lpsltephrdaadflirvft 503

RESULT 14

AAM79025

ID AAM79025 standard; Protein; 702 AA.

AC AAM79025;

DT 06-NOV-2001 (first entry)

DE Human protein SEQ ID NO 1687.

Query Match 17.78; Score 644.5; DB 22; Length 702;
Best Local Similarity 33.78; Pred. NO. 1.7e-49;
Matches 168; Conservative 71; Mismatches 201; Indels 59; Caps 15;

```
QY 12 ELFRDAFPADSSL-FCDDSTPLAOFREDITWRHPOEICATPRLPDDPREGQVKGGL 70
Db 41 elifedplfpaeppslgfkdlg-puskvgnqslswqrpckdlmnpflfmdgispdlcggil 99
QY 71 GDCWFLCACAALOKSRHLDOVIIPGOPSWADQETRGSTFCRIWQFGRWVETFTDDRLPC 130
Db 100 gdcwllanaigslttcprklyrvvprggs--fknyaglfmftqfvgvwnvvvddrlpt 157
QY 131 LAGRLCFSCRCQREDFWFLPLEKVKYAKVHGSYEHLMAQVADALVDTLGGIAERWNLKGV 190
Db 158 knqklvfvhstersefwsallekayaklsgsyaelsgslmeglfdtfgvagsfqlyq-- 215
QY 191 AGSGGQODRPGRWEHFTCRQLHLKDDCLISCCVLSPRAGARE-----LGEFHAFIVSD 244
Db 216 -----ppgnllflllka--verslmgsleivtsdslemtckmlvryhaysvtg 265
QY 245 LRELQAGAGCCILLRIQNPWGRRCWGLWREGEGMSQYDAVAASELSQLOEGEPWVE 304
Db 266 lqdvh-yrgkmetllrvnmpwgrlewngawsdaarewevasdldmqllhktedgetms 324
QY 305 EEEFLREFDELTVG--YPTVEAGHLQSLITERLLCHTRALPGAMVKQSNAGCCRNNSG-F 361
Db 325 ygdflnflleicnltptdlsgdyksyw-----htfyegswirgssagsgcrnpgtlf 378
QY 362 PSNPKFMLRYSE--PSEVYIAVLQSRRLHAADWAGR--ARALVGDGSHTSMSPASIPGK 415
Db 379 wtnpgfkfialpegddped-----daegnvvvctclvalmqknwtharqgga 424
QY 416 HIOAVGLHLMKVKRKRYNLTFRV-LSMPVAVAGTACHAV-----DREVHLRCELSPGYTLA 468
Db 425 qlqtlqfivlyavpketfnqldvnlkkefftkyqdhgfselftnsrevsaglrllppegyl 484
QY 469 VPSTFLKADAPGEFLRVFS 487
Db 485 ipstcphndadtlirvft 503
```

Search completed: September 19, 2002, 14:49:58
Job time: 67 sec

